

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:35:35 ; Search time 34.6897 Seconds
(without alignments)
4429.120 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501
Sequence: 1 caagatctcagatattcaacaa.....tcagtcgaggaacacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	419.6	83.8	1158	US-08-471-724-1	Sequence 1, Appli
2	419.6	83.8	1158	US-08-471-969-1	Sequence 1, Appli
3	419.6	83.8	1158	US-08-384-137-1	Sequence 1, Appli
4	419.6	83.8	1158	US-08-470-006A-1	Sequence 1, Appli
5	419.6	83.8	1158	US-08-691-563C-1	Sequence 1, Appli
6	419.6	83.8	1158	US-09-200-990-1	Sequence 1, Appli
7	419.6	83.8	1158	US-09-133-411-1	Sequence 1, Appli
8	419.6	83.8	2391	US-08-691-563C-57	Sequence 57, Appli
9	404.4	80.7	1537	US-08-691-563C-89	Sequence 89, Appli
10	271.8	54.3	2330	US-09-120-653D-4	Sequence 4, Appli
11	271.8	54.3	3910	US-09-120-653D-1	Sequence 1, Appli
12	212.6	42.4	297	US-08-471-724-2	Sequence 2, Appli
13	212.6	42.4	297	US-08-471-969-2	Sequence 2, Appli
14	212.6	42.4	297	US-08-384-137-2	Sequence 2, Appli
15	212.6	42.4	297	US-08-470-006A-2	Sequence 2, Appli
16	212.6	42.4	297	US-08-691-563C-2	Sequence 2, Appli
17	212.6	42.4	297	US-09-200-990-2	Sequence 2, Appli
18	212.6	42.4	297	US-09-133-411-2	Sequence 2, Appli
19	211.4	42.2	645	US-08-471-724-8	Sequence 8, Appli
20	211.4	42.2	645	US-08-471-969-8	Sequence 8, Appli
21	211.4	42.2	645	US-08-384-137-8	Sequence 8, Appli
22	211.4	42.2	645	US-08-470-006A-8	Sequence 8, Appli
23	211.4	42.2	645	US-08-691-563C-8	Sequence 8, Appli
24	211.4	42.2	645	US-09-200-990-8	Sequence 8, Appli
25	211.4	42.2	645	US-09-133-411-8	Sequence 8, Appli
26	128.8	25.7	299	US-08-691-563C-40	Sequence 40, Appli
27	116.4	23.2	2448	US-08-691-563C-53	Sequence 53, Appli

28	109.8	21.9	741	US-08-471-724-9	Sequence 9, Appli
29	109.8	21.9	741	US-08-471-969-9	Sequence 9, Appli
30	109.8	21.9	741	US-08-384-137-9	Sequence 9, Appli
31	109.8	21.9	741	US-08-470-006A-9	Sequence 9, Appli
32	109.8	21.9	741	US-08-691-563C-9	Sequence 9, Appli
33	109.8	21.9	741	US-09-200-990-9	Sequence 9, Appli
34	109.8	21.9	741	US-09-133-411-9	Sequence 9, Appli
35	109.8	21.9	2389	US-08-691-563C-52	Sequence 52, Appli
36	92.6	18.5	4480	US-09-167-322-12	Sequence 12, Appli
37	86.6	17.3	6363	US-08-929-967-6	Sequence 6, Appli
38	85.8	17.1	1140	US-09-603-185-5	Sequence 5, Appli
39	85.8	17.1	3612	US-09-265-013-3	Sequence 3, Appli
40	85.8	17.1	7308	US-09-011-745-3	Sequence 3, Appli
41	85.8	17.1	7308	US-09-011-745-4	Sequence 4, Appli
42	85.8	17.1	7616	US-09-011-745-2	Sequence 2, Appli
43	85.8	17.1	8202	US-08-258-420-13	Sequence 13, Appli
44	85.8	17.1	8332	US-08-850-961-1	Sequence 1, Appli
45	85.8	17.1	8332	US-09-479-776-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-471-724-1
Sequence 1, Application US/08471724
Patent No. 5800980
GENERAL INFORMATION:
APPLICANT: HERVE PERRON
APPLICANT: FRANCOIS MALLETT
APPLICANT: BERNARD MANDRAND
APPLICANT: FREDERIC BEDIN
APPLICANT: FREDERIC BESME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOLY
TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,724
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-471-724-1
Query Match 83.8%; Score 419.6; DB 1; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;
CY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATGACCAAGCTGTACTAGCCCT 60

Db 61 CAAGAACTAGATTATCAATGAGGCTGTGTCCCTATACCCAGCTGTACCTAACCT 120
Qy 61 TATACCTGCTTTCCCAATACAGAGAGAGAGTGGTTTACAGTCTGTGACCTTCAAG 120
Db 121 TATACAGTCTTTCCCAATACAGAGAGAGAGTGGTTTACAGTCTGTGACCTTCAAG 180
Qy 121 GATGCTCTTCTGTGATCCCTGTACATCTCTGACTCTCAATCTGTGCTTGAAGAT 180
Db 181 GATGCTCTTCTGTGATCCCTGTACATCTCTGACTCTCAATCTGTGCTTGAAGAT 240
Qy 181 ACTTCAACCCCAACATCTCACTCACTGACTATTTTACCCCAAGGTTCAAGGAT 240
Db 241 CTTTGAACCCCAACATCTCACTCACTGACTATTTTACCCCAAGGTTCAAGGAT 300
Qy 241 CCCCATCTATTGGCCAGGATTTAGCCCAAGCTTGAAGCTTCAATCTGTGACCT 298
Db 301 CCCCATCTATTGGCCAGGATTTAGCCCAAGCTTGAAGCTTCAATCTGTGACCT 360
Qy 299 CTTGCTCTTGGTGAAGTATTTACTTTTGGCCGCCATTCAGAACTTGTGCCAT 358
Db 361 CTTGCTCTTGGTGAAGTATTTACTTTTGGCCGCCATTCAGAACTTGTGCCAT 420
Qy 361 CAAGCCACCCAAAGCGCTCTTCAATTTCTGCTGCTACCTGTGCTACATGTTTCCAAACCA 418
Db 421 CAAGCCACCCAAAGCGCTCTTCAATTTCTGCTGCTACCTGTGCTACATGTTTCCAAACCA 480
Qy 419 AAGGCTCAACTGCTGCTCAACAGAG---GTTACTTAAGGCTTAAATATCAAGGACCC 474
Db 481 AAGGCTCAACTGCTGCTCAACAGAGATTAGATCTTAAGGCTTAAATATCAAGGACCC 540
Qy 475 AAGGCTCAACTGCTGCTCAACAGAGATTAGATCTTAAGGCTTAAATATCAAGGACCC 501
Db 541 AAGGCTCAACTGCTGCTCAACAGAGATTAGATCTTAAGGCTTAAATATCAAGGACCC 567

RESULT 2
US-08-471-969-1
; Sequence 1, Application US/08471969
; Patent No. 5871745

GENERAL INFORMATION:
APPLICANT: Hevea PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BESME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESS: Olliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,969
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-471-969-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2,5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

Qy 1 CAAGATCTAGAGATTATCAATGAGGCTGTGTCCCTATACCCAGCTGTACCTAACCT 60
Db 61 CAAGAACTAGAGATTATCAATGAGGCTGTGTCCCTATACCCAGCTGTACCTAACCT 120
Qy 61 TATACCTGCTTTCCCAATACAGAGAGAGAGTGGTTTACAGTCTGTGACCTTCAAG 120
Db 121 TATACAGTCTTTCCCAATACAGAGAGAGAGTGGTTTACAGTCTGTGACCTTCAAG 180
Qy 121 GATGCTCTTCTGTGATCCCTGTACATCTCTGACTCTCAATCTGTGCTTGAAGAT 180
Db 181 GATGCTCTTCTGTGATCCCTGTACATCTCTGACTCTCAATCTGTGCTTGAAGAT 240
Qy 181 ACTTCAACCCCAACATCTCACTCACTGACTATTTTACCCCAAGGTTCAAGGAT 240
Db 241 CTTTGAACCCCAACATCTCACTCACTGACTATTTTACCCCAAGGTTCAAGGAT 300
Qy 241 CCCCATCTATTGGCCAGGATTTAGCCCAAGCTTGAAGCTTCAATCTGTGACCT 298
Db 301 CCCCATCTATTGGCCAGGATTTAGCCCAAGCTTGAAGCTTCAATCTGTGACCT 360
Qy 299 CTTGCTCTTGGTGAAGTATTTACTTTTGGCCGCCATTCAGAACTTGTGCCAT 358
Db 361 CTTGCTCTTGGTGAAGTATTTACTTTTGGCCGCCATTCAGAACTTGTGCCAT 420
Qy 361 CAAGCCACCCAAAGCGCTCTTCAATTTCTGCTGCTACCTGTGCTACATGTTTCCAAACCA 418
Db 421 CAAGCCACCCAAAGCGCTCTTCAATTTCTGCTGCTACCTGTGCTACATGTTTCCAAACCA 480
Qy 419 AAGGCTCAACTGCTGCTCAACAGAG---GTTACTTAAGGCTTAAATATCAAGGACCC 474
Db 481 AAGGCTCAACTGCTGCTCAACAGAGATTAGATCTTAAGGCTTAAATATCAAGGACCC 540
Qy 475 AAGGCTCAACTGCTGCTCAACAGAGATTAGATCTTAAGGCTTAAATATCAAGGACCC 501
Db 541 AAGGCTCAACTGCTGCTCAACAGAGATTAGATCTTAAGGCTTAAATATCAAGGACCC 567

RESULT 3
US-08-384-137-1
; Sequence 1, Application US/08384137
; Patent No. 5871996

GENERAL INFORMATION:
APPLICANT: Hevea PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BESME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESS: Olliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384.137
FILING DATE: February 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-384-137-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140; Indels 6; Gaps 2;
Matches 466; Conservative 0; Mismatches 35;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGACTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATACCTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTCTGATGATCCGTAATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 180
DB 181 GATGCTCTCTGATGATCCGTAATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 240
QY 181 ACTTCAAAACCAACATCTCAACTGACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 240
DB 241 CCTTGAACCCAGACGCTCACTGACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 300
QY 241 CCCCATCTATTGTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 298
DB 301 CCCCATCTATTGTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 299 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 361 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 359 CAAGCACCAGGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGATGATGAT 418
DB 421 CAAGCACCAGGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGATGATGAT 480
QY 419 AAGGCTCACTCTGCTCAAGCAG----GTTACTTAGGGCTAAATTAATCCAAAGGCACC 474
DB 481 AAGGCTCGGCTCTGCTCAAGCAGATTAGATCTAAGGGCTAAATTAATCCAAAGGCACC 540
QY 475 AAGGCTCTCACTGATGAGAAACATCCAG 501
DB 541 AAGGCTCTCACTGATGAGAAACATCCAG 567

RESULT 4
US-08-470-006A-1
Sequence 1, Application US/08470006A
GENERAL INFORMATION:
APPLICANT: Herve PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BEDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND
TITLE OF INVENTION: BIOPOLYMER CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470.006A
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-470-006A-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140; Indels 6; Gaps 2;
Matches 466; Conservative 0; Mismatches 35;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGACTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATACCTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTCTGATGATCCGTAATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 180
DB 181 GATGCTCTCTGATGATCCGTAATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 240
QY 181 ACTTCAAAACCAACATCTCAACTGACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 240
DB 241 CCTTGAACCCAGACGCTCACTGACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 300
QY 241 CCCCATCTATTGTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 298
DB 301 CCCCATCTATTGTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 299 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 361 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 359 CAAGCACCAGGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGATGATGAT 418
DB 421 CAAGCACCAGGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGATGATGAT 480
QY 419 AAGGCTCACTCTGCTCAAGCAG----GTTACTTAGGGCTAAATTAATCCAAAGGCACC 474
DB 481 AAGGCTCGGCTCTGCTCAAGCAGATTAGATCTAAGGGCTAAATTAATCCAAAGGCACC 540
QY 475 AAGGCTCTCACTGATGAGAAACATCCAG 501
DB 541 AAGGCTCTCACTGATGAGAAACATCCAG 567

RESULT 5

US-08-691-563C-1

Sequence 1, Application US/08691563C

Patent No. 6001987

GENERAL INFORMATION:

APPLICANT: HERVE PERRON

APPLICANT: FREDERIC BESEME

APPLICANT: FREDERIC BESIME

APPLICANT: GLAUCIA PARANHOS-BACCALA

APPLICANT: FLORENCE KOMURIAN-PRADEL

APPLICANT: COLETTE JOLIVET

APPLICANT: BERNARD MANDRAND

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC

TITLE OF INVENTION: THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 92

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OLIFF & BERRIDGE

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/691,563C

FILING DATE: 02-AUG-1996

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, William P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 38588

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1158 base pairs

TYPE: nucleotide

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-691-563C-1

Query Match 83.8%; Score 419.6; DB 3; Length 1158;

Best Local Similarity 91.9%; Pred. No. 2.5e-140;

Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTTATCATGAGCTGTTGTTCTCTATACCGACGCTGTACCTGCT 60
DB 61 CAAGAACTCAGGATTTATCATGAGCTGTTGTTCTCTATACCGACGCTGTACCTGCT 120
QY 61 TATATCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTTTCTTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 180
DB 181 GATGCTTTCTTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 240
QY 181 ACTTCAAAACCAATCTTCACTACCTGAGCTATTTTATCCCAAGGCTTCAAGGATAGT 240
DB 241 CTTTGAACCCCAAGCTTCACTACCTGAGCTGTTTATCCCAAGGCTTCAAGGATAGC 300
QY 241 CCCCATCTATTGTCGAGGATTAAGCCCAAGCTTGAAGCATCTCTCACTTCTGAGCA-- 298
DB 301 CCCCATCTATTGTCGAGGATTAAGCCCAAGCTTGAAGCATCTCTCACTTCTGAGCACT 360

QY 299 CTGTGCTTCGAGTGTGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGCCAT 358
DB 361 CTGTGCTTCGAGTGTGATGATTTACTTTAGTCGCCGCCCATTCAGAAACCTTGCCAT 420
QY 359 CAAGCCACCCCAAGGCTTCTTCAATTTCTGCTACCTGTGGCTTCAAGGTTCCAAACCA 418
DB 421 CAAGCCACCCCAAGGCTTCTTCAATTTCTGCTACCTGTGGCTTCAAGGTTCCAAACCA 480
QY 419 AAGGCTCACTGCTCTACAGCAGCAG----GTTACTAGGAGCTTAAATTTCAAGGACAC 474
DB 481 AAGGCTCGGCTCTGCTCTACAGGATTTAGATCTAAGGAGCTTAAATTTCAAGGACAC 540
QY 475 AGGCGCTCTAGTGAAGAACATTCAG 501
DB 541 AGGCGCTCTAGTGAAGAACATTCAG 567

RESULT 6

US-09-200-990-1

Sequence 1, Application US/09200990

Patent No. 6184025

GENERAL INFORMATION:

APPLICANT: HERVE PERRON

APPLICANT: FRANCOIS MALLEET

APPLICANT: BERNARD MANDRAND

APPLICANT: FREDERIC BESIME

APPLICANT: FREDERIC BESIME

TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OLIFF & BERRIDGE

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/200,990

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/471,969

FILING DATE: June 6, 1995

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, William P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 36055A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1158 bases

TYPE: nucleotide

STRANDEDNESS: single-stranded

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-09-200-990-1

Query Match 83.8%; Score 419.6; DB 4; Length 1158;

Best Local Similarity 91.9%; Pred. No. 2.5e-140;

Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTTATCATGAGCTGTTGTTCTCTATACCGACGCTGTACCTGCT 60
DB 61 CAAGAACTCAGGATTTATCATGAGCTGTTGTTCTCTATACCGACGCTGTACCTGCT 120

QY 61 TATACTGCTTTCCCAATAACAGAGAGAGAGTGGTTTAAAGTCTGAGACCTTCAG 120
DB 121 TATACAGTGTCTTCCCAATATACAGAGAGAGAGTGGTTTAAAGTCTGAGACCTTCAG 180
QY 121 GATGCTTCTTCTGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 181 GATGCTTCTTCTGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 181 ACTTCAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 240
DB 241 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 300
QY 241 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 298
DB 301 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 360
QY 299 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 358
DB 361 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 420
QY 359 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 418
DB 421 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 480
QY 419 AAGGCTCACTCTGCTCAACAG---GTTACTTGAAGGCTTAAATTAATCAAGGACAC 474
DB 481 AAGGCTCACTCTGCTCAACAG---GTTACTTGAAGGCTTAAATTAATCAAGGACAC 540
QY 475 AAGGCTCACTCTGCTCAACAG---GTTACTTGAAGGCTTAAATTAATCAAGGACAC 501
DB 541 AAGGCTCACTCTGCTCAACAG---GTTACTTGAAGGCTTAAATTAATCAAGGACAC 567

RESULT 7

US-09-133-411-1
Sequence 1, Application US/09133411
Patent No. 6342383
GENERAL INFORMATION:
APPLICANT: HERVE PERRON
APPLICANT: FRANCOIS MALLETT
APPLICANT: BERNARD MANDRAND
APPLICANT: FREDERIC BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/133,411
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/384,137
FILING DATE: February 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-133-411-1

Query Match 83.8%; Score 419.6; DB 4; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2,5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTAATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGATACCTCT 60
DB 61 CAAGATCTCAGGATTAATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGATACCTCT 120
QY 61 TATACTGCTTTCCCAATAACAGAGAGAGAGTGGTTTAAAGTCTGAGACCTTCAG 120
DB 121 TATACTGCTTTCCCAATAACAGAGAGAGAGTGGTTTAAAGTCTGAGACCTTCAG 180
QY 121 GATGCTTCTTCTGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 181 GATGCTTCTTCTGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 181 ACTTCAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 240
DB 241 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 300
QY 241 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 298
DB 301 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 360
QY 299 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 358
DB 361 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 420
QY 359 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 418
DB 421 CTTTGAACCAACATCTCACTCACTGATGATGATGATGATGATGATGATGATGATGAT 480
QY 419 AAGGCTCACTCTGCTCAACAG---GTTACTTGAAGGCTTAAATTAATCAAGGACAC 474
DB 481 AAGGCTCACTCTGCTCAACAG---GTTACTTGAAGGCTTAAATTAATCAAGGACAC 540
QY 475 AAGGCTCACTCTGCTCAACAG---GTTACTTGAAGGCTTAAATTAATCAAGGACAC 501
DB 541 AAGGCTCACTCTGCTCAACAG---GTTACTTGAAGGCTTAAATTAATCAAGGACAC 567

RESULT 8

US-08-691-563C-57
Sequence 57, Application US/08691563C
Patent No. 6001987
GENERAL INFORMATION:
APPLICANT: HERVE PERRON
APPLICANT: FREDERIC BESEME
APPLICANT: GLAUCIA PARANHOS-BACCALA
APPLICANT: FLORENCE KOMORIAN-PRADEL
APPLICANT: COLETTE JOLIVET
APPLICANT: BERNARD MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 2391 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-691-563C-57

Query Match 83.8%; Score 419.6; DB 3; Length 2391;
Best Local Similarity 91.9%; Pred. No. 3.8e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGAGTATTAATGAGCGTGTGTTCTCTATATAGCCAGCTGTACCTGACCT 60
DB 646 CAAGATCTCAGAGTATTAATGAGCGTGTGTTCTCTATATAGCCAGCTGTACCTGACCT 705
QY 61 TATACTGCTTTCCCAATAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 706 TATACTGCTTTCCCAATAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 121 GATGCTTCTTCTGCATCTCTGATCATCTGATCTCAATTTCTTGTGCTTTGAAGAT 180
DB 766 GATGCTTCTTCTGCATCTCTGATCATCTGATCTCAATTTCTTGTGCTTTGAAGAT 825
QY 181 ACTTGAACCCCAACATCACTCACTGAGTATTTTACCCCAAGGGTTCAGGAGTAT 240
DB 826 CTTTGAACCCCAACATCACTCACTGAGTATTTTACCCCAAGGGTTCAGGAGTAT 885
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGAGCTTGAAGCACTCTCAATCTGAGCA-- 298
DB 886 CCCCATCTATTTGGCCAGGATTAAGCCCAAGAGCTTGAAGCACTCTCAATCTGAGCACT 945
QY 299 CTTGTCTTGGTAGTGTGATTTACTTTTGGCCGCCCATTTCAAGAACTTTGGCCAT 358
DB 946 CTTGTCTTGGTAGTGTGATTTACTTTTGGCCGCCCATTTCAAGAACTTTGGCCAT 1005
QY 359 CAAGCACCCCAAGCGCTCTCAATTTCCGCTACCTGTCGATGATGTTTCCAAACA 418
DB 1006 CAAGCACCCCAAGCGCTCTCAATTTCCGCTACCTGTCGATGATGTTTCCAAACA 1065
QY 419 AAGGCTCACTCTGCTCAAGCAG---GTTACTTAGGGCTAAATTTATCCAAAGGACAC 474
DB 1066 AAGGCTCGGCTGTGCTCAAGCAGGATTAAGTACTTAAAGGCTAAATTTATCCAAAGGACAC 1125
QY 475 AAGGCTCTCAAGTGAAGAAACATCCAG 501
DB 1126 AAGGCTCTCAAGTGAAGAAACATCCAG 1152

RESULT 9
US-08-691-563C-89
Sequence 89, Application US/08691563C
Patent No. 6001987
GENERAL INFORMATION:
APPLICANT: Herve PERRON
APPLICANT: Frederic BESEME
APPLICANT: Frederic BEDIN
APPLICANT: Gaucica PARANHOS-BACCALA

APPLICANT: Florence KOMIRIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ollif & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 1577 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-691-563C-89

Query Match 80.7%; Score 404.4; DB 3; Length 1577;
Best Local Similarity 89.1%; Pred. No. 8.3e-135;
Matches 449; Conservative 0; Mismatches 51; Indels 4; Gaps 1;

QY 1 CAAGATCTCAGAGTATTAATGAGCGTGTGTTCTCTATATAGCCAGCTGTACCTGACCT 60
DB 998 CAAGATCTCAGAGTATTAATGAGCGTGTGTTCTCTATATAGCCAGCTGTACCTGACCT 1057
QY 61 TATACTGCTTTCCCAATAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 1058 TATACTGCTTTCCCAATAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
QY 121 GATGCTTCTTCTGCATCTCTGATCATCTGATCTCAATTTCTTGTGCTTTGAAGAT 180
DB 1118 GATGCTTCTTCTGCATCTCTGATCATCTGATCTCAATTTCTTGTGCTTTGAAGAT 1177
QY 181 ACTTGAACCCCAACATCACTCACTGAGTATTTTACCCCAAGGGTTCAGGAGTAT 240
DB 1178 CTTTGAACCCCAACATCACTCACTGAGTATTTTACCCCAAGGGTTCAGGAGTAT 1237
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGAGCTTGAAGCACTCTCAATCTGAGCACT 300
DB 1238 CCCCATCTATTTGGCCAGGATTAAGCCCAAGAGCTTGAAGCACTCTCAATCTGAGCACT 1297
QY 301 TGTCTTGGTAGTGTGATTTACTTTTGGCCGCCCATTTCAAGAACTTTGTGCATCA 360
DB 1298 TGTCTTGGTAGTGTGATTTACTTTTGGCCGCCCATTTCAAGAACTTTGTGCATCA 1357
QY 361 AGCCACCCCAAGCGCTTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAAGCAAA 420
DB 1358 AGCCACCCCAAGCGCTTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAAGCAAA 1417
QY 421 GGCTCACTGCTCAGCAGCAGGT---TACTTAGGGCTAAATTTATCCAAAGGACACAG 476
DB 1418 GGCTCACTGCTCAGCAGCAGGTAAATTTACTTAGGGCTAAATTTATCCAAAGGACACAG 1477

QY 477 GGCCCTCAGTGAAGAACATCCCA 500
Db 1478 GGCCCTCTGTGAGGAATGATATCCA 1501

RESULT 10

US-09-120-653D-4
; Sequence 4, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120, 653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 4
; LENGTH: 2330
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-4

Query Match 54.3%; Score 271.8; DB 4; Length 2330;
Best Local Similarity 73.8%; Pred. No. 3.7e-87;
Matches 374; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTTCAATGAGGCTGTTGCTCTATAGCCAGCTGTACCTAGCCCT 60
Db 624 CAAGATCTTACATCAATGAGGCTGTTGCTCTATAGCCAGCTGTACCTAGCCCT 683
QY 61 TATATCTGCTTTCCCAATCCAGAGGAGAGGAGGTTTACAGTCTGAGACCTTGA 120
Db 664 TATATCTGCTTTCCCAATCCAGAGGAGAGGAGGTTTACAGTCTGAGACCTTGA 743
QY 121 GATGCTCTCTCTGATCCCTGTACCTGATCTGATCTCAATCTTGTGGCTTTGAAGAT 180
Db 744 GATGCTCTCTCTGATCCCTGTACCTGATCTGATCTCAATCTTGTGGCTTTGAAGAT 803
QY 181 ACTTCAAAACCAATCTCACTCACTGACTATTTTAAACCAAGGTTTCAAGGATAGT 240
Db 804 CCAACAAACCAATCTCACTCACTGACTATTTTAAACCAAGGTTTCAAGGATAGT 863
QY 241 CCCCATCTATTGGGCGAGGATTAACCCCAAGCTTGAGCCATCTCTCA--TACCTGACA 298
Db 864 CCAATCTATTGGGCGAGGATTAACCCCAAGCTTGAGCCATCTCTCA--TACCTGACA 923
QY 299 CTGTCCTTCCGTAGTGGATGATTTTATTTGGCGCCCATTTAGAAACCTTGGCCAT 358
Db 924 CTGTCCTTCCGTAGTGGATGATTTTATTTGGCGCCCATTTAGAAACCTTGGCCAT 983
QY 359 CAAGCCACCAAGCGCTCTTCAATTTCTCGTACCTGCTGCTGATGATGTTTCAAAACA 418
Db 984 CAGGCTACTAGATCTCTTGAACCTTCAAGCTATCAAGGCTAAGAGTGTCTATGTTTC 1043
QY 419 AAGGCTCAACTCTGCTACAGAGGT----TACTTAGGGCTTAAATTTTCAAAAGGACC 474
Db 1044 AAGGCTCAACTCTGCTACAGAGGTAAATTTTCAAGGCTTAAATTTTCAAAAGGACA 1103
QY 475 AGGCGCTCAGTGAAGAACATCCAG 501
Db 1104 AGGCGCTCAGTGAAGAACATCCAG 1130

RESULT 11

US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120, 653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 1
; LENGTH: 3910
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

Query Match 54.3%; Score 271.8; DB 4; Length 3910;
Best Local Similarity 73.8%; Pred. No. 5e-87;
Matches 374; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTTCAATGAGGCTGTTGCTCTATAGCCAGCTGTACCTAGCCCT 60
Db 1020 CAAGATCTTACATCAATGAGGCTGTTGCTCTATAGCCAGCTGTACCTAGCCCT 1079
QY 61 TATATCTGCTTTCCCAATCCAGAGGAGAGGAGGTTTACAGTCTGAGACCTTGA 120
Db 1080 TATATCTGCTTTCCCAATCCAGAGGAGAGGAGGTTTACAGTCTGAGACCTTGA 1139
QY 121 GATGCTCTCTCTGATCCCTGTACCTGATCTGATCTCAATCTTGTGGCTTTGAAGAT 180
Db 1140 GATGCTCTCTCTGATCCCTGTACCTGATCTGATCTCAATCTTGTGGCTTTGAAGAT 1199
QY 181 ACTTCAAAACCAATCTCACTCACTGACTATTTTAAACCAAGGTTTCAAGGATAGT 240
Db 1200 CCAACAAACCAATCTCACTCACTGACTATTTTAAACCAAGGTTTCAAGGATAGT 1259
QY 241 CCCCATCTATTGGGCGAGGATTAACCCCAAGCTTGAGCCATCTCTCA--TACCTGACA 298
Db 1260 CCAATCTATTGGGCGAGGATTAACCCCAAGCTTGAGCCATCTCTCA--TACCTGACA 1319
QY 299 CTGTCCTTCCGTAGTGGATGATTTTATTTGGCGCCCATTTAGAAACCTTGGCCAT 358
Db 1320 CTGTCCTTCCGTAGTGGATGATTTTATTTGGCGCCCATTTAGAAACCTTGGCCAT 1379
QY 359 CAAGCCACCAAGCGCTCTTCAATTTCTCGTACCTGCTGCTGATGATGTTTCAAAACA 418
Db 1380 CAGGCTACTAGATCTCTTGAACCTTCAAGCTATCAAGGCTAAGAGTGTCTATGTTTC 1439
QY 419 AAGGCTCAACTCTGCTACAGAGGT----TACTTAGGGCTTAAATTTTCAAAAGGACC 474
Db 1440 AAGGCTCAACTCTGCTACAGAGGTAAATTTTCAAGGCTTAAATTTTCAAAAGGACA 1499
QY 475 AGGCGCTCAGTGAAGAACATCCAG 501
Db 1500 AGGCGCTCAGTGAAGAACATCCAG 1526

RESULT 12

US-08-471-724-2
; Sequence 2, Application US/08471724
; Patent No. 5800980
; GENERAL INFORMATION:
; APPLICANT: HEVE PERRON
; APPLICANT: FRANCOIS WALLENT
; APPLICANT: BERNARD MANDRAND

```

; APPLICANT: Frederic BEDIN
; APPLICANT: Frederic BESEME
; TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
; TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
; TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,724
; FILING DATE: June 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36055C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 bases
; TYPE: nucleotide
; STRANDEDNESS: single-stranded
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
;
US-08-471-724-2
;
Query Match 42.4%; Score 212.6; DB 1; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
QY 1 CAGATCTCAGGATTATCATAGAGCTGTGTTCTCTATACCGAGCTGTACTAGCCCT 60
DB 61 CAGAGACTCAGGATTATCATAGAGCTGTGTTCTCTATACCGAGCTGTACTAGCCCT 120
QY 61 TATACAGTCTTCCCAATAATCCAGAGAGAGAGTGTATACAGTCTGTGACCTTAG 120
DB 121 TATACAGTCTTCCCAATAATCCAGAGAGAGAGTGTATACAGTCTGTGACCTTAG 180
QY 121 GATGCTTTCTTCCCAATAATCCAGAGAGAGAGTGTATACAGTCTGTGACCTTAG 180
DB 181 GATGCTTTCTTCCCAATAATCCAGAGAGAGAGTGTATACAGTCTGTGACCTTAG 240
QY 181 ACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAACGCTCACTCACTGACTGTTTACCCCAAGGTTCAAGG 295
;
RESULT 13
US-08-471-969-2
; Sequence 2, Application US/08471969
; Patent No. 5871745
; GENERAL INFORMATION:
; APPLICANT: Hervé PERRON
; APPLICANT: Francois MALLET
; APPLICANT: Bernard MANDRAND
; APPLICANT: Frederic BEDIN
; APPLICANT: Frederic BESEME
; TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
; TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
; TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
;

```

```

; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,969
; FILING DATE: June 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36055A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 bases
; TYPE: nucleotide
; STRANDEDNESS: single-stranded
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
;
US-08-471-969-2
;
Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
;
QY 1 CAGATCTCAGGATTATCATAGAGCTGTGTTCTCTATACCGAGCTGTACTAGCCCT 60
DB 61 CAGAGACTCAGGATTATCATAGAGCTGTGTTCTCTATACCGAGCTGTACTAGCCCT 120
QY 61 TATACAGTCTTCCCAATAATCCAGAGAGAGAGTGTATACAGTCTGTGACCTTAG 120
DB 121 TATACAGTCTTCCCAATAATCCAGAGAGAGAGTGTATACAGTCTGTGACCTTAG 180
QY 121 GATGCTTTCTTCCCAATAATCCAGAGAGAGAGTGTATACAGTCTGTGACCTTAG 180
DB 181 GATGCTTTCTTCCCAATAATCCAGAGAGAGAGTGTATACAGTCTGTGACCTTAG 240
QY 181 ACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAACGCTCACTCACTGACTGTTTACCCCAAGGTTCAAGG 295
;
RESULT 14
US-08-384-137-2
; Sequence 2, Application US/08384137
; Patent No. 5871996
; GENERAL INFORMATION:
; APPLICANT: Hervé PERRON
; APPLICANT: Francois MALLET
; APPLICANT: Bernard MANDRAND
; APPLICANT: Frederic BEDIN
; APPLICANT: Frederic BESEME
; TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
; TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
; TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
;

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,137
FILING DATE: February 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-384-137-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAAGACTCAGGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATACTGCTTTTCCAAATACAGAGAGACAGAGTGTTCAGTCTGAGCTTCAAG 120
DB 121 TATACAGTGTCTTCCAAATACAGAGAGACAGAGTGTTCAGTCTGAGCTTCAAG 180
QY 121 GATGCTTTCTTGCAATCCCTGTACATCTGACTCTCAATTTCTTGTGCTTTGAAGAT 180
DB 181 GATGCTTTCTTGCAATCCCTGTACATCTGACTCTCAATTTCTTGTGCTTTGAAGAT 240
QY 181 ACTTCAACCCCAACATCTCACTGACCTGACATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAACGCTCAACTCACTGACCTGACATTTTACCCCAAGGTTCAAGG 295

RESULT 15

US-08-470-006A-2
Sequence 2, Application US/08470006A
Patent No. 5962217
GENERAL INFORMATION:
APPLICANT: Heave PERRON
APPLICANT: Francois MALLEF
BEST APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND
TITLE OF INVENTION: BIOPOLYMER CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,006A
FILING DATE: June 6, 1995

ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-470-006A-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAAGACTCAGGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATACTGCTTTTCCAAATACAGAGAGACAGAGTGTTCAGTCTGAGCTTCAAG 120
DB 121 TATACAGTGTCTTCCAAATACAGAGAGACAGAGTGTTCAGTCTGAGCTTCAAG 180
QY 121 GATGCTTTCTTGCAATCCCTGTACATCTGACTCTCAATTTCTTGTGCTTTGAAGAT 180
DB 181 GATGCTTTCTTGCAATCCCTGTACATCTGACTCTCAATTTCTTGTGCTTTGAAGAT 240
QY 181 ACTTCAACCCCAACATCTCACTGACCTGACATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAACGCTCAACTCACTGACCTGACATTTTACCCCAAGGTTCAAGG 295

Search completed: April 17, 2003, 07:38:02
Job time: 39.6897 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 17, 2003, 04:57:20 ; Search time 123.415 Seconds
(without alignments)
4077.979 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501

Sequence: 1 caagatccagcagatcatca.....tcagtcgagacacacacag 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications NA:*

2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

3: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	431.8	86.2	1600	7	US-08-979-847-207 Sequence 207, App
2	425.6	85.0	1600	7	US-08-979-847-206 Sequence 206, App
3	422.2	84.3	2304	7	US-08-979-847-87 Sequence 87, App1
4	419.6	83.8	1158	7	US-08-979-847-1 Sequence 1, App1
5	419.6	83.8	1597	7	US-08-979-847-205 Sequence 205, App
6	419.6	83.8	2365	7	US-08-979-847-88 Sequence 88, App1
7	419.6	83.8	2391	7	US-08-979-847-53 Sequence 53, App1
8	404.4	80.7	1577	7	US-08-979-847-83 Sequence 83, App1
9	402.2	80.3	568	10	US-09-864-761-16537 Sequence 16537, A
10	236.6	47.2	557	10	US-09-864-761-11891 Sequence 11891, A
11	231.6	46.2	449	10	US-09-864-761-2921 Sequence 2921, Ap
12	221.6	44.2	467	10	US-09-864-761-1194 Sequence 1194, Ap
13	212.6	42.4	297	7	US-08-979-847-2 Sequence 2, App1
14	212	42.3	540	10	US-09-864-761-15266 Sequence 15266, A
15	211.4	42.2	645	7	US-08-979-847-8 Sequence 8, App1
16	182.6	36.4	438	7	US-08-979-847-129 Sequence 129, App
17	182.6	36.4	438	7	US-08-979-847-130 Sequence 130, App
18	182.2	36.4	438	7	US-08-979-847-128 Sequence 128, App
19	182.2	36.4	438	7	US-08-979-847-131 Sequence 131, App

20	181.8	36.3	438	7	US-08-979-847-201 Sequence 201, App
21	165.6	33.1	944	10	US-09-864-761-14911 Sequence 14911, A
22	146.6	29.3	429	7	US-08-979-847-138 Sequence 138, App
23	145.4	29.0	429	7	US-08-979-847-135 Sequence 135, App
24	145	28.9	429	7	US-08-979-847-136 Sequence 136, App
25	145	28.9	429	7	US-08-979-847-137 Sequence 137, App
26	137.2	27.4	150	7	US-08-979-847-36 Sequence 36, App1
27	116.4	23.2	2448	7	US-08-979-847-49 Sequence 49, App1
28	114.2	22.8	191	10	US-09-864-761-19702 Sequence 19702, A
29	109.8	21.9	741	7	US-08-979-847-9 Sequence 9, App1
30	109.8	21.9	2389	7	US-08-979-847-48 Sequence 48, App1
31	107.6	21.5	133	10	US-09-864-761-33021 Sequence 33021, A
32	104.4	20.8	431	10	US-09-864-761-28100 Sequence 28100, A
33	102.2	20.4	181	10	US-09-864-761-17958 Sequence 17958, A
34	92.6	18.5	447	10	US-09-864-761-11512 Sequence 11512, A
35	92.4	18.4	326014	10	US-09-731-231A-3 Sequence 3, App1
36	89	17.8	157	10	US-09-864-761-31788 Sequence 31788, A
37	85.8	17.1	2151	10	US-09-845-157-1 Sequence 1, App1
38	85.8	17.1	8332	10	US-09-006-298-1 Sequence 1, App1
39	82.8	16.5	143068	10	US-09-967-768A-316 Sequence 2, App1
40	77.8	15.5	8323	10	US-09-970-597-2 Sequence 28561, A
41	76.8	15.3	183	10	US-09-864-761-28561 Sequence 28561, A
42	76	15.2	685	10	US-09-864-761-31430 Sequence 31430, A
43	70.8	14.1	1436	10	US-09-864-761-17414 Sequence 17414, A
44	69.8	13.9	85	7	US-08-979-847-195 Sequence 195, App
45	69.4	13.9	593	10	US-09-864-761-7065 Sequence 7065, App

ALIGNMENTS

RESULT 1

US-08-979-847-207

Sequence 207, Application US/08979847

Publication No. US2003039664A1

GENERAL INFORMATION:

APPLICANT: PERRON, HERVE

APPLICANT: BESEME, FREDERIC

APPLICANT: BEDIN, FREDERIC

APPLICANT: PARANHOS-BACCALA, GLAUCIA

APPLICANT: KOMRINOS-PRADTEL, FLORENCE

APPLICANT: JOLIVET-REINAUD, COLETTE

APPLICANT: MANDRAND, BERNARD

APPLICANT: GARSON, JEREMY

APPLICANT: TUXE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC

TITLE OF INVENTION: THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC

STREET: P.O. BOX 19928

CITY: ALEXANDRIA

STATE: VA

COUNTRY: USA

ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,847

FILING DATE: 26-NOV-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: MPB 39046A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 207:

SEQUENCE CHARACTERISTICS:
LENGTH: 1600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-979-847-207

Query Match 86.2%; Score 431.8; DB 7; Length 1600;
Best Local Similarity 90.5%; Pred. No. 5,5e-133;
Matches 459; Conservative 20; Mismatches 22; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCATGAGGCTGTGTCTCTATATACGACGCTGTACCTTACCTT 60
DB 373 CAAGAATCAGGATTATCATGAGGCTGTGTCTCTATATACGACGCTGTACCTTACCTT 432
QY 61 TATACCTGCTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 433 TATACAGGCTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
QY 121 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCT 180
DB 493 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCT 552
QY 181 ACTTCAAAACCAACATCTTCACTCACTGATCTTATACCCCAAGGCTTCAAGGATAGT 240
DB 553 CTTTGAACCCCAACGCTTCACTCACTGATCTTATACCCCAAGGCTTCAAGGATAGT 612
QY 241 CCCCATCTATTGGGCGAGGATTTAGCCCAAGCTTGAAGCAATCTCTATACCTTGACCA-- 298
DB 613 CCCCATCTATTGGGCGAGGATTTAGCCCAAGCTTGAAGCAATCTCTATACCTTGACCA 672
QY 299 CTTCGCTTCTGATGATGATTTATTTTGGCGCCCATTTGAAGAACTTGTGACAT 358
DB 673 CTTCGCTTCTGATGATGATTTATTTTGGCGCCCATTTGTGACATTTGTGACAT 732
QY 359 CAAGCAACCAAGCGCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCA 418
DB 723 CAAGCAACCAAGCGCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCA 792
QY 419 AAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTAAATTTATCCAAAGGACAC 474
DB 793 AAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTAAATTTATCCAAAGGACAC 852
QY 475 AGGGCCCTCAGTGAAGAACATCCAG 501
DB 853 AGGGCCCTCAGTGAAGAACATCCAG 879

RESULT 2

US-08-979-847-206
Sequence 206, Application US/08979847
Publication No. US2003003964A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIEF & BERRIDGE, PLC
STREET: P O BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA

ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 206:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-979-847-206

Query Match 85.0%; Score 425.6; DB 7; Length 1600;
Best Local Similarity 88.6%; Pred. No. 6,3e-131;
Matches 449; Conservative 26; Mismatches 24; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCATGAGGCTGTGTCTCTATATACGACGCTGTACCTTACCTT 60
DB 373 CAAGAATCAGGATTATCATGAGGCTGTGTCTCTATATACGACGCTGTACCTTACCTT 432
QY 61 TATACCTGCTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 433 TATACAGGCTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
QY 121 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCT 180
DB 493 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCT 552
QY 181 ACTTCAAAACCAACATCTTCACTCACTGATCTTATACCCCAAGGCTTCAAGGATAGT 240
DB 553 CTTTGAACCCCAACGCTTCACTCACTGATCTTATACCCCAAGGCTTCAAGGATAGT 612
QY 241 CCCCATCTATTGGGCGAGGATTTAGCCCAAGCTTGAAGCAATCTCTATACCTTGACCA-- 298
DB 613 CCCCATCTATTGGGCGAGGATTTAGCCCAAGCTTGAAGCAATCTCTATACCTTGACCA 672
QY 299 CTTCGCTTCTGATGATGATTTATTTTGGCGCCCATTTGAAGAACTTGTGACAT 358
DB 673 CTTCGCTTCTGATGATGATTTATTTTGGCGCCCATTTGTGACATTTGTGACAT 732
QY 359 CAAGCAACCAAGCGCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCA 418
DB 723 CAAGCAACCAAGCGCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCA 792
QY 419 AAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTAAATTTATCCAAAGGACAC 474
DB 793 AAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTAAATTTATCCAAAGGACAC 852
QY 475 AGGGCCCTCAGTGAAGAACATCCAG 501
DB 853 AGGGCCCTCAGTGAAGAACATCCAG 879

RESULT 3

US-08-979-847-87
Sequence 87, Application US/08979847
Publication No. US2003003964A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE

APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKU, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 87:
SEQUENCE CHARACTERISTICS:
LENGTH: 2304 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
US-08-979-847-87

Query Match 84.3%; Score 422.2; DB 7; Length 2304;
Best Local Similarity 92.3%; Pred. No. 9.9e-130;
Matches 468; Conservative 0; Mismatches 33; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTAATAGCCAGCTGTACCTAGCCCT 60
DB 640 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTAATAGCCAGCTGTACCTAGCCCT 699
QY 61 TATATCTGCTTCCCAATAATCAGAGAGAGAGGTGTTACAGTCTTGACCTTCAAG 120
DB 700 TATATCTGCTTCCCAATAATCAGAGAGAGAGGTGTTACAGTCTTGACCTTCAAG 759
QY 121 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 760 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 819
QY 181 ACTTCAAAACCAACATCTCAATCACTGAGCTATTTTACCCCAAGGTTGAGGATAGT 240
DB 820 CTTTGAACCAACATCTCAATCACTGAGCTATTTTACCCCAAGGTTGAGGATAGT 879
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTGAGCAATCTCTCATCTGAGCA-- 298
DB 880 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTGAGCAATCTCTCATCTGAGCACT 939
QY 299 CTGTGCTTGGTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 940 CTGTGCTTGGTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 999
QY 359 CAAGCACCACCAAGCGCTTCAATTTCTGCTGATGAGTGAATGTTTCCAAACCA 418

DB 1000 CAAGCACCACCAAGCTTCTTACTTCTCTACTACTGCTGCTCAAGGTTTCCAAACCA 1059
QY 419 AAGCTCAACTCTGCTCAACAGCAG-----GTTACTTAGGGCTAAATATTCAAAGGACCC 474
DB 1060 AAGCTGAGCTCTGCTCAACAGGATTAAGATTAAGGCTTAATATATCAAAAGGACCC 1119
QY 475 AGGGCTCTAGTAGGAACACATCCAG 501
DB 1120 AGGGCTCTAGTAGGAACACATCCAG 1146

RESULT 4
US-08-979-847-1
Sequence 1, Application US/08979847
Publication No. US20030039664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKU, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-1

Query Match 83.8%; Score 419.6; DB 7; Length 1158;
Best Local Similarity 91.9%; Pred. No. 5.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTAATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTAATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATATCTGCTTCCCAATAATCAGAGAGAGAGGTGTTACAGTCTTGACCTTCAAG 120
DB 121 TATATCTGCTTCCCAATAATCAGAGAGAGAGGTGTTACAGTCTTGACCTTCAAG 180

QY 121 GATGCTTCTTCTGATCCCTGTACATCTGATCTCTCAATTCCTGTTTCCCTTTGAAGAT 180
DB 181 GATGCTTCTTCTGATCCCTGTACATCTGATCTCTCAATTCCTGTTTCCCTTTGAAGAT 240
QY 181 ACTTCAAAACCAACATCTCAATCTGACCTGTGACTATTTTACCCCAAGGTTCCAGGATAGT 240
DB 241 CCTTTGAACCCCAACCTCTCAATCTGACCTGTGACTATTTTACCCCAAGGTTCCAGGATAGC 300
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCCATTCCTTACCTGTGACA-- 298
DB 301 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCATTTCTTACCTGTGACA-- 360
QY 299 CTGTCTCTTGGTGTAGTGTATTTACTTTTGTGACCCCATCTCAAGAACTTTGTGCAAT 358
DB 361 CTGTCTCTTGGTGTAGTGTATTTACTTTTGTGACCCCATCTCAAGAACTTTGTGCAAT 420
QY 359 CAAGCACCACCAAGGCTCTTCAATTTCTGCTGCTTCTGTGGTTACATGTTTCCAAACCA 418
DB 421 CAAGCACCACCAAGGCTCTTCAATTTCTGCTGCTTCTGTGGTTACATGTTTCCAAACCA 480
QY 419 AAGGCTCAACTGCTGCTCAGACGAG---GTACTTAGGGCTAAATTTATCCAAAGGACAC 474
DB 481 AAGGCTCGGCTTGTCTCAGACGAGATTTAGATCTTAGGGCTAAATTTATCCAAAGGACAC 540
QY 475 AAGGCTCTCAGTGTAGGAACACATCCAG 501
DB 541 AAGGCTCTCAGTGTAGGAACGATTCAG 567

RESULT 5

US-08-979-847-205

; Sequence 205, Application US/08979847

; Publication No. US20030039664A1

; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
APPLICANT: BESENE, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIVIER & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPA 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 205:
SEQUENCE CHARACTERISTICS:
LENGTH: 1597 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-979-847-205

Query Match 83.8%; Score 419.6; DB 7; Length 1597;
Best Local Similarity 91.9%; Pred. No. 6.1e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCCAGGATTTATCAATAGAGCTGTGTCTCTATATGACAGGCTGTACAGCCCT 60
DB 373 CAAGATCCAGGATTTATCAATAGAGCTGTGTCTCTATATGACAGGCTGTACAGCCCT 432
QY 61 TATACAGTCTTTCCCAATATCCAGAGAACAGAGTGTATTAAGTCTTGAAGCTTCAAG 120
DB 433 TATACAGTCTTTCCCAATATCCAGAGAACAGAGTGTATTAAGTCTTGAAGCTTCAAG 492
QY 121 GATGCTTCTTCTGATCCCTGTACATCTGATCTCTCAATTCCTGTTTCCCTTTGAAGAT 180
DB 493 GATGCTTCTTCTGATCCCTGTACATCTGATCTCTCAATTCCTGTTTCCCTTTGAAGAT 552
QY 181 ACTTCAAAACCAACATCTCAATCTGACCTGTGACTATTTTACCCCAAGGTTCCAGGATAGT 240
DB 553 CTTTGAACCCCAACGCTCAACTCACTGACTGTTTACCCCAAGGTTCCAGGATAGC 612
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCCATCTTACCTGTGACA-- 298
DB 613 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCCATCTTACCTGTGACA-- 672
QY 299 CTGTCTCTTGGTGTAGTGTATTTACTTTTGGCCCATCTCAAGAACTTTGTGCAAT 358
DB 673 CTGTCTCTTGGTGTAGTGTATTTACTTTTGGCCCATCTCAAGAACTTTGTGCAAT 732
QY 359 CAAGCACCACCAAGGCTCTTCAATTTCTGCTGCTTCTGTGGTTACATGTTTCCAAACCA 418
DB 733 CAAGCACCACCAAGGCTCTTCAATTTCTGCTGCTTCTGTGGTTACATGTTTCCAAACCA 792
QY 419 AAGGCTCAACTGCTGCTCAGACGAG---GTACTTAGGGCTAAATTTATCCAAAGGACAC 474
DB 793 AAGGCTCGGCTTGTCTCAGACGAGATTTAGATCTTAGGGCTAAATTTATCCAAAGGACAC 852
QY 475 AAGGCTCTCAGTGTAGGAACACATCCAG 501
DB 853 AAGGCTCTCAGTGTAGGAACGATTCAG 879

RESULT 6

US-08-979-847-88

; Sequence 88, Application US/08979847

; Publication No. US20030039664A1

; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
APPLICANT: BESENE, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIVIER & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-2787
TELEFAX: 703-836-6400
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 2365 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
US-08-979-847-88

Query Match 83.8%; Score 419.6; DB 7; Length 2365;
Best Local Similarity 91.9%; Pred. No. 7.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGAGTATTAATGAGGCTGTGCTCTATAGCCGCTGTACCTGACCT 60
DB 646 CAAGAACTCAGAGTATTAATGAGGCTGTGCTCTATAGCCGCTGTACCTGACCT 705
QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 706 TATACAGTCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 121 GATGCTCTTCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCT 180
DB 766 GATGCTCTTCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCT 825
QY 181 ACTTCAAAACCAATCTCACTGACCTGACCTGACCTGACCTGACCTGACCTGAC 240
DB 826 CTTTGAACCCCAAGCTCTCACTGACCTGACCTGACCTGACCTGACCTGACCTG 885
QY 241 CCCCATCTATTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 298
DB 886 CCCCATCTATTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 945
QY 299 CTTGCTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 358
DB 946 CTTGCTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1005
QY 359 CAAGCCACCAAGGCTCTTCAATTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418
DB 1006 CAAGCCACCAAGGCTCTTCAATTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1065
QY 419 AAGGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 474
DB 1066 AAGGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1125
QY 475 AAGGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 501
DB 1126 AAGGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1152

RESULT 7
US-08-979-847-53
Sequence 53, Application US/08979847
Publication No. US20030039664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESENE, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA

APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GABSON, JEREMY
APPLICANT: TUXE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIVIER & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-2787
TELEFAX: 703-836-6400
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 2391 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-53

Query Match 83.8%; Score 419.6; DB 7; Length 2391;
Best Local Similarity 91.9%; Pred. No. 7.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGAGTATTAATGAGGCTGTGCTCTATAGCCGCTGTACCTGACCT 60
DB 646 CAAGAACTCAGAGTATTAATGAGGCTGTGCTCTATAGCCGCTGTACCTGACCT 705
QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 706 TATACAGTCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 121 GATGCTCTTCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCT 180
DB 766 GATGCTCTTCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCTGTCATCT 825
QY 181 ACTTCAAAACCAATCTCACTGACCTGACCTGACCTGACCTGACCTGACCTGAC 240
DB 826 CTTTGAACCCCAAGCTCTCACTGACCTGACCTGACCTGACCTGACCTGACCTG 885
QY 241 CCCCATCTATTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 298
DB 886 CCCCATCTATTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 945
QY 299 CTTGCTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 358
DB 946 CTTGCTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1005
QY 359 CAAGCCACCAAGGCTCTTCAATTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418
DB 1006 CAAGCCACCAAGGCTCTTCAATTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1065

Qy 419 AAGGCTCACTCTGCTCAACAG---GTTACTTAGGGCTAAAATTATCCAAAGGACC 474
Db 1066 AAGGCTCGCTCTGCTCAACAGAGATTAGATCTTAGGGCTAAATATCCAAAGGACC 1125
Qy 475 AGGGCCCTCAGTGAAGAACATCCAG 501
Db 1126 AGGGCCCTCAGTGAAGAACATCCAG 1152

RESULT 8

US-08-979-847-83
Sequence 83, Application US/08979847
Publication No. US2003003964A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES;
NUMBER OF INVENTIONS: 210
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIEF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 1577 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-979-847-83

Query Match 80.7%; Score 404.4; DB 7; Length 1577;
Best Local Similarity 89.1%; Pred. No. 6,7e-124;
Matches 449; Conservative 0; Mismatches 51; Indels 4; Gaps 1;

Qy 1 CAAGATCTCAGATTATCATGAGGCTGTTCTCTATAGCAAGCTGTACTTACCCCT 60
Db 998 CAAGATCTCAGATTATCATGAGGCTGTTCTCTATAGCAAGCTGTACTTACCCCT 1057
Qy 61 TATACTCTGCTTCCCAATACCAAGAGAGAGTGTATTACAGTCTTGACCTTCAG 120
Db 1058 TATACTCTGCTTCCCAATACCAAGAGAGAGTGTATTACAGTCTTGACCTTCAG 1117
Qy 121 GATGCTCTTCTGCATCCCTGTACATCTGACTGCAATTCCTGTTGCTTTGAAGAT 180

Db 1118 GATGCTCTTCTGCATCCCTGTACATCTGATTTCTCAATTCCTGTTGCTTTGAAGAT 1177
Qy 181 ACTTCAAACCCCAATCTCAACTCAGCTGAGACTATTTTACCCCAAGGTTCCAGGATAGT 240
Db 1178 CTTTGAACCCCAATGTTCTCAATTCACCTGAGACTGTTTACCCCAAGGTTCCAGGATAGC 1237
Qy 241 CCCCATCTATTGGCCAGGCAATTGCCCCAAGACTTGAAGCCCAATCTCAATACCTTGACACT 300
Db 1238 CCCCATCTATTGGCCAGGCAATTGCCCCAAGACTTGAAGCCCAATCTCAATACCTTGACACT 1297
Qy 301 TGTCTTCGGTAGGTGATGATTTACTTTTGGCCGCCAATTCAGAAACCTGTGCCATCA 360
Db 1298 TTGTCCTTCGGTAGGTGATGATTTTACTTTTGGCCGCCAATTCAGAAACCTGTGCCATCA 1357
Qy 361 AGCCACCAAGCGCTCTTCATTTCTCGTACTGCTGCTACATGCTTTCCAAACCAA 420
Db 1358 AGCCACCAAGCGCTCTTCATTTCTCGTACTGCTGCTACATGCTTTCCAAACCAA 1417
Qy 421 GGCTCACTCTGCTCAGCAGGCT---TACTTGGGCTAAATTTATCCAAAGGACCCAG 476
Db 1418 GGCTCAGCTCTGCTCAGCAGGCTTAAATCTTAGGGTTAAATTTATCCAAAGGACCCAG 1477
Qy 477 GGCCCTCAGTGAAGAACATCCCA 500
Db 1478 GGCCCTCAGTGAAGAACATCCCA 1501

RESULT 9

US-09-864-761-16537
Sequence 16537, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO: 16537
LENGTH: 568
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: MAP TO AC007567.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
US-09-864-761-16537

Query Match 80.3%; Score 402.2; DB 10; Length 568;
Best Local Similarity 91.1%; Pred. No. 2.3e-122; Indels 7; Gaps 3;
Matches 462; Conservative 0; Mismatches 38;

QY 1 CAAGATCTCAGGATTATCATGAGGCTGTTCTCTTATACCACTGTAACCTTACCTT 60
DB 58 CAAGATCTCAGGATTATCATGAGGCTGTTCTCTTATACCACTGTAACCTTACCTT 117
QY 61 TATACCTGCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 118 TATACCTGCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177
QY 121 GATGCTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCT 180
DB 178 GATGCTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCT 237
QY 181 ACTTGAACCCCAATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 240
DB 238 CCTTGAACCCCAATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 297
QY 241 CCCCATCTATTGTCGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 298
DB 298 CCCCATCTATTGTCGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 357
QY 299 CTGTCCTCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 358 CTGTCCTCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417
QY 359 CAAGCAGCCAGAGGCTCTTCAATTTCTGCTACCTGCTGCTACATGCTGCTGCTGCTGCT 418
DB 418 CAAGCAGCCAGAGGCTCTTCAATTTCTGCTACCTGCTGCTACATGCTGCTGCTGCTGCT 477
QY 419 AAGGCTCACTGCTGCTCAAGAGG---TTACTTAGGGCTAAATTTATCCAAAGGCAAC 474
DB 478 AA-GGCTCACTGCTGCTCAAGAGG---TTACTTAGGGCTAAATTTATCCAAAGGCAAC 536
QY 475 AAGGCTCTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 501
DB 537 AAGGCTCTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 563

RESULT 10
US-09-864-761-11981

Sequence 11981, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO: 11981
LENGTH: 557
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007923.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.73
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-11981

Query Match 47.2%; Score 236.6; DB 10; Length 557;
Best Local Similarity 72.2%; Pred. No. 2e-68;
Matches 337; Conservative 0; Mismatches 124; Indels 6; Gaps 2;

QY 41 AGCCAGCTGTAAGTCCCTTATCTGCTTTCCCAATATCCAGAGGAGAGAGTGT 100
DB 1 ACCCAGTTATACCCCAACCCCTGTAACCTGCTCTCAAAATACCAAGAGGAGAGATGAT 60
QY 101 TTACAGTCTGAGACCTTAGAGATGCTTCTGATGATGCTGATGATGCTGATGATGAT 160
DB 61 TCACGTCTGAGACCTCAAGAGATGCTTCTGATGATGCTGATGATGCTGATGATGAT 120
QY 161 TCTTGTGCTTGAAGATATCTTCAACCAACCAATCTCACTCACTGACTATTTTAC 220
DB 121 TTCTCTTGTGCTTGAAGATATCTTCAACCAACCAATCTCACTCACTGACTATTTTAC 180
QY 221 CCCAAGGCTCAGAGATATGCTCCCATCTATTGTCGAGGAGATGAGCCCAAGATTTAGCC 280
DB 181 CCCAAGGCTTAAAGGATATGCTCCCATCTATTGTCGAGGAGATGAGCCCAAGATTTAGCC 240
QY 281 AATCTCA--TACCTGACACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338
DB 241 ACTTCAAGTCCAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300

QY	ATTGGAACCTGTGGCATTAAGCAACCAAG-GGCTTCGAATTCCTGGTACCTGTG	398
Db	301 GTTTGGAAGCCTCATGTGACGAGGCTACTTCAATCTTGAACTTTCGTAGTATTAAG	360
QY	399 GCTACATGGTTCCTCCAAACCAAGGGCTCAACTGT-ATCAGACAGGT-----TACTTAGGGC	454
Db	361 GGTACAGGCACTCTTAGGTGGAAGGCTCAGCTTTATCTACAGCAGGTCAAAATATCTAGGCC	420
QY	455 TAAATTTATCCAAAGGACACAGGGGCCCTCAGTGAAGAAACATGCTAG	501
Db	421 TAAATTTATCCAGAGGACACAGGGGCCCTCAGCAAKGAATGATATCAG	467

```

RESULT 11
US-09-864-761-2921/c
: Sequence 2921, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aecmica-X-1
: CURRENT FILING DATE: 2001-05-23
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,166
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine ver11.1
: SEQ ID NO 2921
: LENGTH: 449
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC010856.2
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3

```

1	OTHER INFORMATION:	EXPRESSED IN HEART, SIGNAL = 1.7
2	OTHER INFORMATION:	EXPRESSED IN ADULT LIVER, SIGNAL = 4.6
3	OTHER INFORMATION:	EXPRESSED IN BT474, SIGNAL = 2.7
4	OTHER INFORMATION:	EXPRESSED IN BONE MARROW, SIGNAL = 1.7
5	OTHER INFORMATION:	EXPRESSED IN PLACENTA, SIGNAL = 2.6
6	OTHER INFORMATION:	EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
7	OTHER INFORMATION:	EXPRESSED IN BRAIN, SIGNAL = 1.5
8	OTHER INFORMATION:	EXPRESSED IN LUNG, SIGNAL = 3
9	US-09-864 -761-2921	

Oy	41	AGCAGGTGACCTACACCCCTTATCTGCTTCCCAAATACGAGGAAAGCAGACTGGT	10
Db	449	ATCCAGGTATACCCAAACCCCTTATACCTGCTCTCMAAATACGAGGAAAGCAGATGGT	39
Oy	101	TTACAGTCTCGACCTTCAGAGATGCTCTTCTGATCCTGTATCCTGACTCTCAAT	16
Db	389	TCATATGTTCTGGACCTCAAGATGCTTCCCTGTGATTCCTCGACCTCGACTCCAAAG	33
Oy	161	TCCTGTTGCTTGAAGTATCTTCAAAACCAACATCTCAACTACCTGACTATTTTAC	22
Db	329	TTCTCTTGGCTTTGAGGATCCCAACAGCACAGTCCCACTTTGCTGATGCTTGC	27
Oy	221	CCCAAGGGTTCAGGGATATCTCCCATCTATTGGCCACAGCATTAAGCCACATTAGGC	28
Db	269	CTCAAGGGTTTAGGATAGCTTCAATCTGTTTGGTCAGGCACTGGCCCAAGATCTAGGCC	21
Oy	281	AATCTCTA--TACCTGGAACCTTGTCCCTCGGTAGGTGATGATTAATCTTTGGCCGCC	33
Db	209	ACTTCTCAAGTACAGGACCTGTGCTCTTCAGTATGAGAGATTTATCTATGTTACA	15
Oy	339	ATTCAAGAACTTGTGCTCATCAAGCACCCAGGCGCTTCAATTTCTCGTACCTGTG	39
Db	149	GTTCAAAAGCTCATATCCACAGGCTAATCTAGCTCTTTGAAATTTCTATGCTAATCAAG	90
Oy	399	GCTACATGTTTCCAAACCAAGGCTCAACTCTGTCTAC-AGCAGTTACTTAGGCTTAA	45
Db	89	GGTACATGAGCATTAATCAAGAGGCCACAGCTCTGCTTACAAATATCTAGGCTTAA	30
Oy	458	AATTATCCAAAGGCAACAGGCGCTCAG	485
Db	29	TCCTTAGCCAAAGGAAACCAAGGCTTAAAG	2

```

RESULT 12
US-09-864-761-1194/C
; Sequence 1194 Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30

```

PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 1194
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006999.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN B1474, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.82
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
US-09-864-761-1194

Query Match 44.2%; Score 221.6; DB 10; Length 467;
Best Local Similarity 70.7%; Pred. No. 1.8e-63;
Matches 325; Conservative 0; Mismatches 129; Indels 6; Gaps 2;

48 TGTACTAGCCCTTATCTGCTTCCCAATACAGAGAGAGAGAGTGTATACAGT 107
Db 467 TGCACCCAAACCTTATCCCTGCTCTCAAAATACAGAGAGATGATGTTATGTT 408
QY 108 CCTGACCTTGAAGATGCTTCTTGTGATCCCTGATCTGATCTCAATCTTGT 167
Db 407 TCTGACCTCAAGATATCTTCTTATTCCTGCACTGACTCCCAATTTCTCTT 348
QY 168 TGCCTTGAAGATCTTAAACCCCAATCACTCACTGAGATATTTTACCCCAAG 227
Db 347 TGCCTTGAAGATCTTAAACCCCAATCACTCACTGAGATATTTTACCCCAAG 288
QY 228 GTTCAGGATAGTCCCATCTTATTTGGCCAGGATTAAGCCCAAGCTTACCATCTC 287
Db 287 GTTAAAGGATAGTCCCATCTTATTTGGCCAGGATTAAGCCCAAGCTTACCATCTC 228
QY 288 ATACCTGG--ACACTTGTCTTGGTGAAGTGAATATTTACTTTGGCCGCCCATTTACA 345
Db 227 AAATCCACCACTCTGCTCTTCAATATGTGATATTTACTTTGGCTACAGATTTCA 168
QY 346 AACCTTGGCATCAAGCACCAGGCGCTTCAATTTCCGCTACCTGTGGCTATCAT 405
Db 167 AGCTTCATGCAAGCGGCTACTTGAATCTCTTGAATTTCTAGTAAATCAAGGGTACAA 108
QY 406 GGTTCACCAACCAAGGCTCACTGCTCAACAGAGG---TTACTTAAAGGCTTAAAT 461

Db 107 GCTGTTAGTCAAGGCCCGAGCTTGTGCTTACAGAGGCCCAATATCTAGCTTATCAT 48
QY 462 ATCCAAAGCACAGGCGCCCTCAGTGAGGAACATCATCAG 501
Db 47 AGCCAGAGGACCAAGGCCCTCAGCAAGGAATGAAAAAG 8

RESULT 13
US-08-979-847-2
Sequence 2, Application US/08979847
Publication No. US2003039664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BODIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUXE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIVIER BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-2

Query Match 42.4%; Score 212.6; DB 7; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.4e-60;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

1 CAAGATCGAGGATTAATCAATGAGGCTGTTCTCTATAGCAGGCTGATACCTAGCCCT 60
Db 61 CAAGATCGAGGATTAATCAATGAGGCTGTTCTCTATAGCAGGCTGATACCTAGCCCT 120
QY 61 TATACCTGCTTCCCAATACAGAGAAAGCAGAGTGTATTAAGTCTGAGACCTTACAG 120
Db 121 TATACAGTGTCTTCCCAATACAGAGAAAGCAGAGTGTATTAAGTCTGAGACCTTACAG 180
QY 121 GATGCTTCTTGTGATCTGATACCTGATACCTGATCTCAATTTCTTGTGCTTGAAGAT 180
Db 181 GATGCTTCTTGTGATCTGATACCTGATACCTGATCTCAATTTCTTGTGCTTGAAGAT 240


```

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 645 base pairs
;   TYPE: nucleotide
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-979-847-8

```

```

Query Match      42.2%; Score 211.4; DB 7; Length 645;
Best Local Similarity 89.9%; Pred. No. 5e-60;
Matches 250; Conservative 0; Mismatches 22; Indels 6; Gaps 2;

```

```

QY 230 TCAGGATAGTCCCATCTATTGGCCAGGATTAGCCCAAGACTTGAAGCCATCTCAT 289
Db 1 TCAGGATAGCCCCCATCTATTGGCCAGGATTAGCCCAAGACTTGAAGCCATCTCAT 60
QY 290 ACCTGGACA--CTGTCTTGGGTAGGTGATGATTACTTTGGCCGCCATTGAGAAA 347
Db 61 ACCTGGACACTCTGTCTTGGGTAGGTGATGATTACTTTAGTCCGCCGTTGAGAAA 120
QY 348 CCTGTGCTATCAAGGCCAAGGCGTCTTCAATTTGCTGCTACTGTGGCTACATGG 407
Db 121 CCTGTGCTATCAAGGCCAAGGCGTCTTCAATTTGCTGCTACTGTGGCTACATGG 180
QY 408 TTTCCAAACCAAAGGCTCACTGTGCTCAGCAG---GTTACTTAGGGCTAAATTAT 463
Db 181 TTTCCAAACCAAAGGCTGTGGCTGTGCTCAGAGGATAGATACTTAGGGCTAAATTAT 240
QY 464 CCAAGGCAACCAAGGCGCTCAGTGAAGAAACATCCAG 501
Db 241 CCAAGGCAACCAAGGCGCTCAGTGAAGAAACATCCAG 278

```

Search completed: April 17, 2003, 07:46:44
 Job time : 127.415 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:30:25 ; Search time 1268.84 seconds

(without alignments)
6394.748 Million cell updates/secTitle: US-09-719-554-3_COPY_5000_5500
Sequence: 1 caagatctcagatcatca.....tcagtgaggaacacatccag 501Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	384.2	76.7	521	17	AZ517418 RPCI-11-4
2	349	69.7	674	17	AQ111469 CIT-HSP-2
3	324.2	64.7	533	17	B17809 34719.TVB
4	300	59.9	594	17	AQ193544 CIT-HSP-2
5	266.8	53.3	595	17	B59017 CIT-HSP-201
6	265.6	53.0	745	9	AU121943 AU121943

C 7	256.6	51.2	750	9	AL698185
C 8	253.2	50.5	231	12	BF989696
C 10	252.6	50.4	707	17	B66771
C 11	208	41.5	736	10	AV731083
C 12	206.2	41.2	537	17	AQ270891
C 13	204.4	40.8	416	17	AQ041889
C 14	204.4	40.8	518	17	AQ361037
C 15	203.8	40.7	434	10	BE168410
C 16	201	40.1	408	17	AQ033042
C 17	190	37.9	549	17	AQ231407
C 18	177.8	35.5	405	17	AQ033015
C 19	175.8	35.4	557	17	AQ061550
C 20	168	33.5	676	9	AL705035
C 21	164.8	32.9	518	17	AQ243044
C 22	161.6	32.3	442	17	AQ132081
C 23	159.2	31.8	541	17	AQ075734
C 24	157.2	31.4	491	17	AQ070222
C 25	155	30.9	431	17	B17135
C 26	152.4	30.4	442	17	AQ221075
C 27	150.6	30.1	467	17	AQ0701392
C 28	148	29.5	456	17	AQ0876467
C 29	148	29.5	519	17	AQ232635
C 30	147.8	29.5	526	17	AQ153050
C 31	146	29.1	449	17	AQ0701255
C 32	143	28.5	532	17	AQ217989
C 33	140.6	28.1	737	17	AQ024733
C 34	140.4	28.0	437	17	B67137
C 35	140.2	27.6	461	17	B47870
C 36	138.4	26.6	518	17	AQ225099
C 37	132.6	26.5	518	17	AQ0807540
C 38	132.2	26.4	529	17	AQ284525
C 39	131.4	26.2	429	17	AQ028930
C 40	127.2	25.4	431	17	CIT-HSP-201
C 41	123	24.6	469	17	AQ145803
C 42	121	24.2	398	17	AQ223144
C 43	119.8	23.9	454	17	AQ044697
C 44	118.2	23.6	533	17	AQ226226
C 45			383	17	AQ479087
					AQ192721
					HS_2251_B

ALIGNMENTS

RESULT 1
AZ517418/c 521 bp DNA linear GSS 16-OCT-2000
LOCUS RPCI-11-4708.TV RPCI-11 Homo sapiens genomic clone RPCI-11-4708,
DEFINITION DNA sequence.

ACCESSION AZ517418
VERSION AZ517418.1 GI:10826891
KEYWORDS GSS.

SOURCE

human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 521)

AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter

TITLES

JOURNAL BAC end sequences of library RPCI-11

COMMENT Unpublished (1997)
Other GSSs: RPCI11-4708.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genetic cs (info@regen.com). BAC end search page: http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.

Seq primer: T7
Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .521

```

/organism="Homo sapiens"
/db_xref="GDB:7518007"
/db_xref="taxon:9606"
/clone="RPC1-11.4708"
/clone_1b="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/ncbi_vector="pBAC3.6";
Rccl1 Human Male BAC Library
BASE COUNT      145 a      93 c      147 g      136 t
ORIGIN

```

Query Match 76.7%; Score 384.2; DB 17; Length 521;

Best Local Similarity 90.8%; Pred. No. 1e-111;
Matches 433; Conservative 0; Mismatches 38; Indels 6; Gaps 2;

Qy	31	GTGCTCTAATGACAGCTGAACCTGAGCCCTTATACTGCTTTCCCAATACGAGGAA	90
Db	521	GTTCCTTAACCCAGAGCTGACCTTAACCTTACTCTGCTTTCCCAATACGAGGCA	46
Qy	91	GCAGAGTGTTTACAGTCTCGAAGCTTCAGATGCTTCCTTGACATCTCT	15
Db	461	GCAAGTGTTTACAGTCTGACCTTAAGGGTGCTTTCTGTATCTCCGTACATCTCT	40
Qy	151	GACTCTCAATTCCTGTTTGCTTTGAAGTACTTCAAACCAATATCTCAATCACTCG	21
Db	401	GACTCTACTCTTGTTTGCTTTGAAGTCTTGAACCAATATCAATCAATCACTCG	34
Qy	211	ACTATTTTACCCCAAGGTTCAAGAGATAGTCCCATCTATTGGCCAGCATTAGCCAA	27
Db	341	ACTGTTTATCCCAAGGTTCAAGAGATAGTCCCATCTATTGGCCAGCATTAGCCAA	28
Qy	271	GACTTGAACCAATCTCTCATCTGTGACA--CTTGCTCTGGTAGTGATATTACTT	32
Db	281	GACTTGAACCAATCTCTCATCTGTGACAATCTTGCTCTTGAGTAGTGATATTACTT	22
Qy	329	TTGCGCGCCATTCAGAAACCTGTGSCATCAAGCCACCAAGGCGCTTCAATTCCTC	38
Db	221	TTGCGCACCCGTTCAAGAACTTGTGCCATCAAGCCACCAAGCATCTTAAATTTCTTC	16
Qy	389	GCTTACTGTGGCTACATGTTGTTCCAAACCAAGGCTCAACTGTCTCAGACAGGT---	44
Db	161	GCTTACTGTGGTTTACAAAGGTTTCCAAATGAAGGCTCACTCTCTCAGACAGTTAAA	10
Qy	445	TACTTAGGGCTAAATTTATTCAAAGGCACAGAGGCTCTAATGAGAAACATTCAG	501
Db	101	TACTTAGGGCTAAATTTATTCAAAGGCACAGAGGCTCTAATGAGAAATCATTCAG	45

RESULT 2	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AO111469/c	AO111469	674 bp DNA linear	AO111469	AO111469	GI:3488126	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	Adams, M.D., Rounsley, S. D., Zhao S., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.

TITLE	use of a random human BAC End Sequence Database for Sequence-Ready
JOURNAL	Map Building
COMMENT	Unpublished (1998)
	Other_GSSs: CIT-HSP-2372020.TF

FEATURES	Location/Qualifiers
source	1. .674

BASE COUNT	180 a	128 c	184 g	182 t
ORIGIN				

Query Match	69.7%	Score 349	DB 17	Length 674
Best Local Similarity	90.3%	Prod. No. 2.1e-100		
Matches 411	Conservative 0	Mismatches 30	Indels 14	Gaps 3

OY	53	CTAGCCCTTAATACCTCTGCTTTCCCAAAATACAGAGGAAGAGAGTGGTTTACAGTCTGG	11.2
Db	674	CTACCCCTTAATACCTGTGCTTTCCAAATACAGAGGAAGAGAGTGGTTTACAGTCTGG	61.5
OY	113	ACCTTCAGAGATGCGCTTCTTGATGATCCCTGTACATCCTGACTCAATCTTGTTGGCT	17.2
Db	614	ACCTTAAGAGATGCGCTTCTTGATGATCCCTGTGACTCTTGACTCAATCTTAATTTGGCT	55.5
OY	173	TTGGAAGATATTCAAACCCACATCTCACTACCTGAGCATTTTATCCCAAGGGTTCA	23.3
Db	554	TTGAAGATCTTGCAGACCAATATCTCAACTCACTGACTGTTTATCCCAAGGGTTCA	49.5
OY	233	GGGATAGTCCCCATCTAATTTGGCCAGGACATTAGCCCAAGATTGAGCCAATCTCATACC	29.2
Db	434	GGGATAGCCCCCATCTAATTTGGCCAGGACATTAGCCCAAGATTGAGCCAATCTCATACC	43.5
OY	293	TGGACA--CTTGCTCTTGGGTAGGTGATGATTTACTTTTGCGGCCCATTTCAAAAAGCT	35.5
Db	434	AGGACACTCTTGCTCTTCAAGTACATGATGATTTACTTTTAAAGCCGCTGTTCACAAAAGCT	37.5
OY	351	TGTGCATCAAGCCACCCAAAGCGCTCTTCAATTTCTCGCTACCTGTGGCTACATGCTTT	41.0
Db	374	TGTGCATCAAGCCACCCGAGCGCTTTAAATTTCTTGCCAACTG-----TGGTTTT	32.2
OY	411	CCAAACCAAGGCTCAACTGTGCTCAGCAGGT---TACTTAGGGCTAAATATTATCCA	46.5
Db	332	CCAAACCAAGGCTCAGCTGTGCTCAGCAGGTTAAATATTAGGGCTAAATATTATCCA	26.3
OY	467	AAGGACACGAGGCGCTCAGTAGAGGAACATCTCAG	50.1
Db	262	AAGGATCCAGGCGCTCAGTAGAGGAATGTATTCAG	22.8

[illegible]

KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 533)
AUTHORS Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.
TITLE Use of a BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A-347L09"
/clone_lib="CIT978SKA1"
/sex="Female"
/cell_type="Fibroblast"
/note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
Caltech Human BAC Library A1"
BASE COUNT 154 a 97 c 144 g 138 t
ORIGIN

Query Match 64.7%; Score 324.2; DB 17; Length 533;
Best Local Similarity 86.0%; Pred. No. 1.6e-92;
Matches 398; Conservative 0; Mismatches 53; Indels 12; Gaps 3;

QY 50 TACCTAGCCCTTATCTCGCTTCCCAATACGAGGAGGAGGCTTACAGTCC 109
DB 533 TACTTAACCCCTTATACCTGCTTCCCAATGCGAGGAGGAGGCTTATAGTCT 474
QY 110 TGGACCTTCAGGATGCTTCTTCTGCATCCCTGTACATCCCTCAATCTTTGTTG 169
DB 473 TGGACCTTAAAGATGCTTCTTCTGCATCCCTGTACATCCCTCAATCTTTGTTG 414
QY 170 CCTTGAAGATCTTCAACCCCAATCTCACTGACCTGACATTTTACCCCAAGGT 229
DB 413 CCTTGAAGATCTTCAACCCCAATCTCACTGACCTGACATTTTACCCCAAGGT 354
QY 230 TCAGGATATGTCCTCATTTTGGCCAGGCAT-----TAGCCCAAGCTTGAAGCAT 283
DB 353 TCAGGATATGTCCTCATTTTGGCCAGGCATCAACCCCAAGCTTGAAGCAT 294
QY 284 CCTTACCTTCGACACT--TGCTTCGGTAGAGGATATTTACTTTGGCCGCAATT 341
DB 293 TCTCATATCTGAACACTATGTTGTTGTTGATAGATATTTAATCTAGCCGCCCTG 234
QY 342 CAGAAACCTTGTCATCAAGCCAGCCAGCGCTCTTCAATTTCTCGTACCTTGAGCT 401
DB 233 CAGAAACCTTGTCATCAAGCCAGCCAGCGCTTTTAACTTCTCACCACCTTGAGCT 174
QY 402 ACATGTTTCCAAACCAAGGCTCACTCTGCTCAGAGAGT----TACTTAGGCTTAA 457
DB 173 ACATGTTTCCAAACCAAGGCTCACTCTGCTTAAAGAGTTAAATCTTAGGCTTAA 114
QY 458 AATTATCCAAAGCAGGAGGCTCAGTAGGAGAAACATCCA 500
DB 113 AATTATCCAAAGCAGGAGGCTCAGTAGGAGATGTATCCA 71

RESULT 4
A0193544/c

LOCUS A0193544 594 bp DNA linear GSS 16-SEP-1998
DEFINITION CIT-HSP-2381P7.TR CIT-HSP Homo sapiens genomic clone 2381P7, DNA
sequence.
ACCESSION A0193544
VERSION A0193544.1 GI:3603659
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 594)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2381P7.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
source
1..594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2381P7"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelBAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 151 a 111 c 170 g 162 t
ORIGIN

Query Match 59.9%; Score 300; DB 17; Length 594;
Best Local Similarity 91.7%; Pred. No. 9.4e-85;
Matches 341; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

QY 136 ATCCCTGTACATCTGACTCTCAATCTTGTGCTTGAAGATCTTCAACCCACA 195
DB 594 ATCCCTGTACATCTGACTCTCAATCTTGTGCTTGAAGATCTTCAACCTAATG 535
QY 196 TCTCACTCAGCTGAGACTATTTTACCCCAAGGCTTCAAGGATAGTCCCATCTATTGGC 255
DB 534 TCTCACTCAGCTGAGACTATTTTACCCCAAGGCTTCAAGGATAGTCCCATCTATTGGC 475
QY 256 CAGGATATGCGCAAGACTTGAAGCAATCTCATCTGAGACA--CTGTCTTGAGTAG 313
DB 474 CAGGATATGCGCAAGACTTGAAGCAATCTCATCTGAGACACTTGTCTTGAGTAG 415
QY 314 GTGATGATTTACTTTTGGCCGCCCATTCAGAAACCTTGTGCATCAAGCCCAAGCG 373
DB 414 ATGATGATTTACTTTTGGCCGCCCATTCAGAAACCTTGTGCATCAAGCCCAAGCG 355
QY 374 CTCTTCAATTTCTCTGCTCACTGTGCTCACTAGTGTTCGAAACCAAGGCTCAATCTGCG 433
DB 354 CTCTTCAATTTCTCTGCTCACTGTGCTCACTAGTGTTCGAAACCAAGGCTCAATCTGCG 295
QY 434 TCACAGCAGGT----TACTTAGGCTTAAATTTATCCAAAGGACAGGAGGCTCAGTAG 489
DB 294 TCACAGCAGGTTAAATTTACTTAGGCTTAAATTTATCCAAAGGACAGGAGGCTCAGTAG 235
QY 490 GAACACATCCAG 501

Db 234 GAACGTTCCAG 223

RESULT 5
B59017/c 595 bp DNA linear GSS 20-JUN-1998
LOCUS CIT-HSP-2014K19.TR CIT-HSP Homo sapiens genomic clone 2014K19, DNA
DEFINITION sequence.
ACCESSION B59017
VERSION B59017.1 GI:2613735
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 595)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Base,S., Linher,K., Golden
,K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL Unpublished (1997)
COMMENT Other GSSs: CIT-HSP-2014K19.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Classes: BAC ends.

FEATURES
source 1..595
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7042821"
/db_xref="taxon:9606"
/clone="2014K19"
/clone_1lb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 158 a 104 c 159 g 174 t
ORIGIN

Query Match 53.3%; Score 266.8; DB 17; Length 595;
Best Local Similarity 89.2%; Pred. No. 3.9e-74;
Matches 323; Conservative 0; Mismatches 33; Indels 7; Gaps 3;

Qy 146 ATCTGACTCTCAATCTTGTGCTTGGATGAGATCTCAAAACCAACATCTCACTCA 205
Db 595 ATCCGAGCTCTCAATCTTGTGCTTGGATGAGATCTCAAAACCAACATCTCACTCA 536
Qy 206 CCGAGCTATTTACCCCAAGGATTGAGGATAGTCCCATCTATTTGGCCAGGATTAG 265
Db 535 CCGAGTGTGTTACCCCAAGGATTGAGGATAGTCCCATCTATTTGGCCAGGATTAG 476
Qy 266 CCAAGACTTGAAGCAATCTCATACCTGAGACT--TTCCTTGCGTAGTGATGATT 323
Db 475 CTCAAGACTTGAAGCAAGTCTCATACCTGAGACTCTCTTCAAGACATGATGATT 416
Qy 334 TACTTTGGCCGCCATTCAGAACTTGTGCGATCAAC CCAACCAAGCGCTTCTCAATT 383
Db 415 TACTTTAGCTCCCTTTGAGAACTTGTGCGATCAAC CCAACCAAGCATTTTAATT 356
Qy 384 TCCTGCTTACCTGTGGCTACATGTTTCCAAACCAAGGCTCAACTGCTGCTACAGCAGG 443
Db 355 TCCTCTCACCTGTG-TATTAAGGTTTCCAAACCAAGGATGAGCTGCTGCTACAGTAGG 297

Qy 444 T----TACTTAGGGCTAAATATTCAAAGGACACAGGCGCTTCAGTAGGAACATCC 499
Db 296 TTAATACTTAGTGTCTAAATATTCAAAGGACACAGGCGCTTCAGTAGGAACATATCC 237

Qy 500 AG 501
Db 236 AG 235

RESULT 6
AUI21943 745 bp mRNA linear EST 01-AUG-2002
LOCUS AUI21943 MAMMA1 Homo sapiens cDNA clone MAMMA1.001324 5', mRNA
DEFINITION sequence.
ACCESSION AUI21943
VERSION AUI21943.1 GI:10937178
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 745)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isoagui,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isoagui
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES
source 1..745
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMA1.001324"
/clone_1lb="MAMMA1"
/tissue_type="mammary gland"
/note="Vector: pME18SFL3"

BASE COUNT 214 a 196 c 156 g 176 t 3 others
ORIGIN

Query Match 53.0%; Score 265.6; DB 9; Length 745;
Best Local Similarity 89.8%; Pred. No. 1.1e-73;
Matches 309; Conservative 0; Mismatches 29; Indels 6; Gaps 2;

Qy 164 TGTGTCCTTTGAAGATCTTCAAAACCAACATCTCAACTCACTGAGACTATTTTACCC 223
Db 1 TGTGTCCTTTGAAGATCTTCAAAACCAACATCTCAACTCACTGAGACTATTTTACCC 60
Qy 224 AAGGTTGAGGATAGTCCCATCTATTTGGCCAGGATTAGGCCAAGCTTGAAGCCAT 283
Db 61 AAGGTTGAGGATAGTCCCATCTATTTGGCCAGGATTAGGCCAAGCTTGAAGCCAT 120
Qy 284 CCTATACCTGAGACA--CTTGCTTCGTTGAGTAGATGATTTTACTTTGGCCGCCATT 341
Db 121 TCTACACCTGAGACACTTGTCTTGAGTACATGATGATTTACTTTTACCTGCCCTT 180
Qy 342 CAGAAACCTTGTGCGATCAAGCAACCAAGCGCTTCAATTCTCTGCTACCTGTGCT 401
Db 181 CAGAAACCTTGTGCGATCAAGCAACCAAGCGCTTCAATTCTCTGCTACCTGTGCT 240
Qy 402 ACATGTTTCCAAACCAAGGCTCACTGCTGCTACAGCAGG----TTACTTAGGGCTAA 457
Db 241 ACCAGTTTCCAAACCAAGGCTCACTGCTGCTACAGCAGGCTTAATCTTAGGGCTAA 300

QY 458 AATTATCCAAAGCAGCAGGCGCTGAGTGAAGAACATCAG 501
Db 301 AATTATCTTAAGCAGCAGGCGCTGAGTGAAGAACATCAG 344

RESULT 7
LOCUS AL698185 750 bp mRNA linear EST 21-MAR-2002
DEFINITION DKFZP686N20106.r1.686 (synonym: hlc3) Homo sapiens cDNA clone
DKFZP686N20106.5, mRNA sequence.
ACCESSION AL698185
VERSION AL698185.1 GI:19618725
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 750)
AUTHORS Ottenwaelder, B., Obermaier, B., Mewes, W., Mewes, H.W., Weill, B. and Wiemann, S.
TITLE EST (Ottenwaelder, B., Obermaier, B., Mewes, H.W., Weill, B. and Wiemann, S.)
JOURNAL Unpublished (2001)
COMMENT Contact: Ottenwaelder B
MIPS
Am Klopferstr. 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
Sequenced by MediGenomix (Martinsried/Germany) within the CDNA
sequencing consortium of the German Genome Project. No si sequence
available.
This clone (DKFZP686N20106) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubenerweg 6, 14059
Berlin-Charlottenburg; GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
FEATURES
source
1..750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP686N20106"
/clone_lib="686 (synonym: hlc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex; Site_1: SfiI; Site_2: SfiI; CDNA-collection"

BASE COUNT 195 a 151 c 194 g 210 t

ORIGIN
Query Match 51.2%; Score 256.6; DB 9; Length 750;
Best Local Similarity 73.1%; Pred. No. 8.2e-71;
Matches 372; Conservative 0; Mismatches 129; Indels 8; Gaps 3;

QY 1 CAAGATCTCAGATTTATCATGAGGCTGTCTCTCTATAGCCAGTGAAGT--AGCC 58
Db 582 CAAGATCTTAACTATCATGAGGAGTAAATCTCTATATGAGCTGTACCCCAACC 523

QY 59 CTTATATCTCTGCTTTCCCAATATCCAGAGAGAGAGTGTTCAGTCTCGACCTTC 118
Db 522 CTTATATCTCTGCTTTCCCAATATCCAGAGAGAGAGTGTTCAGTCTCGACCTTC 463

QY 119 AGGATGCTCTTCTGATCCCTGATACCTGATCTCAATCTTCTGTTGGCTTTGAG 178
Db 462 AGGATGCTCTTCTGATCCCTGATACCTGATCTCAATCTTCTGTTGGCTTTGAG 403

QY 179 ATACTTCAACCCCAATCTCACTCACTGATCTATTTTACCCCAAGGTTCCAGGATA 238
Db 402 ATCCACAGACCAACGATCTCACTCACTGATCTATTTTACCCCAAGGTTCCAGGATA 343

QY 239 GTCCCATCTATTTGGCCAGGATTAAGCCCAAGATTCCTCA--TACTTGA 296
Db 342 GCCCTAATCTGTGGTGAAGGACATGGCCCAAGATCTAGGCCACTTCAAGTCCAGCA 283

QY 297 CACTGTCTCTGCTGATGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGTGCC 356
Db 282 CTGTGTCTCTGATGATGATGATTTACTTTGGCTTACAGATTCAGAAACCTTGTGCC 223

QY 357 ATCAAGCAGCCAGCCCTCTTCAATTTCTGCTACTGCTGATGATGATTTTCCAAAC 416
Db 222 AGCAGGCTACTCTGATGATGATGATTTCTGATGATGATGATGATGATGATGAT 163

QY 417 CAAGGCTCACTGCTGCTCAAGCAGGAT---TACTTGGGCTAAATATTCAGAGCA 472
Db 162 TGAAGGCCAGGCTGCTGCTTACACAGATCAATATCTAGGCTTAATCTTACGACGAGA 103

QY 473 CCAGGCGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 501
Db 102 CCAACACCTCAGCAAGAAATGAGACAG 74

RESULT 8
LOCUS BF989696/c 291 bp mRNA linear EST 23-JAN-2001
DEFINITION M2-GN0127-231000-004-e11 GN0127 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF989696
VERSION BF989696.1 GI:12396021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 291)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&t2=MR2-GN0127-231000-004-e11&t3=2000-10-23&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 290.
Location/Qualifiers
FEATURES
source
1..291
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0127"
/dev_stage="Adult"
/note="Organ: placenta, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORSTS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 86 a 54 c 82 g 68 t

ORIGIN
Query Match 50.5%; Score 253.2; DB 12; Length 291;
Best Local Similarity 96.4%; Pred. No. 6.1e-70;
Matches 269; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 4 GATCTCAGATTATCATATGAGGCTGTGTTCTCTATAGCCAGCTGTACTAGCCCTTAT 63
Db 280 GAACCTAGGATATATCATAGAGGCTGTGTTCTCTATAGCCAGCTGTACTAGCCCTTAT 221
Qy 64 ACTGCTCTTCCCAATATCCAG 123
Db 220 ACTGCTCTTCCCAATATCCAG 161
Qy 124 GCGCTCTCTGTCAGCTCCGTATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 183
Db 160 GCGCTCTCTGTCAGCTCCGTATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101
Qy 184 TCAATCCCAACATCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243
Db 100 TCAATCCCAACATCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 41
Qy 244 CATCTATTGGCGAGGAGCATTAAGCCCAAGCTTGAAGCA 282
Db 40 CATCTATTGGCGAGGAGCATTAAGCCCAAGCTTGAAGCA 3

RESULT 9

LOCUS B66771 707 bp DNA linear GSS 21-JUN-1998
DEFINITION CIT-HSP-2015D21.TFB CIT-HSP Homo sapiens genomic clone 2015D21, DNA
sequence.
ACCESSION B66771
VERSION B66771.1 GI:2640749
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Base,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL Unpublished (1997)
COMMENT Other-GSS: CIT-HSP-2015D21.TR
CONTACT: Mark Adams
DEPARTMENT: Department of Eukaryotic Genomics
INSTITUTION: The Institute for Genomic Research
ADDRESS: 9712 Medical Center Dr., Rockville, MD 20850, USA
TEL: 301 838 0200
FAX: 301 838 0208
EMAIL: mdadams@tigr.org
CLONES: Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search.html
Seq primer: M13-21
CLASS: BAC ends.

FEATURES

source 1..707
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7043039"
/db_xref="taxon:9606"
/clone="2015D21"
/clone_id="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; site_1: HindIII; site_2:
HindIII"

BASE COUNT 195 a 183 c 144 g 185 t
ORIGIN

Query Match 50.4%; Score 252.6; 1B 17; Length 707;
Best Local Similarity 71.4%; Pred. No. 1.5e-69;
Matches 362; Conservative 0; Mismatches 139; Indels 6; Gaps 2;
Qy 1 CAAGATCTCAGATTATCATATGAGGCTGTGTTCTCTATAGCCAGCTGTACTAGCCCT 60
|||||

Db 176 CAAGATCTCAGATTATCATATGAGGCTGTGTTCTCTATATGCCAGCTGTATCCCAACCC 235
Qy 61 TATATCTGCTTTTCCCAATATCCAG 120
Db 236 TATATCTGCTTTTCCCAATATCCAG 295
Qy 121 GATGCTCTTCTGTCAGCTCCGTATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Db 296 GATGCTCTTCTGTCAGCTCCGTATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 355
Qy 181 ACTTCAACCAACATCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Db 356 CCGGAGAACCCATCTCAATCTTACATGACGGCTCTGCTCAAGGCTTACAGAGATAC 415
Qy 241 CCGCATCTATTGGCGAGGAGCATTAAGCCCAAGCTTGAAGCAATCTCA -TACCTGAGCA 298
Db 416 CCGCATCTATTGGCGAGGAGCATTAAGCCCAAGCTTGAAGCAATCTCA -TACCTGAGCA 475
Qy 299 CTGTGCTCTGCGTAGGTGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGCAT 358
Db 476 CTGATGCTCTGCGTAGGTGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGCAT 535
Qy 359 CAAGCCACCAAGGCTCTTCAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
Db 536 CAGGCTACTCTGATCTCTTGAACCTTCTAGCTATGAGAGGTCAGAGGCTCTCAATCA 595
Qy 419 AAGGCTCACTCTGCTCTCAGCAGAGT---TACTTGAAGGCTTAAATATCCAAAGCACC 474
Db 596 AAGGCCAGCTCTGCTCTCAGCAGAGT---TACTTGAAGGCTTAAATATCCAAAGCAGGCC 655
Qy 475 AAGGCCCTCAGTGAAGAACATCTCAG 501
Db 656 ATGGCTTCATCAAGAGAAAGATATCAG 682

RESULT 10

LOCUS AV731083 736 bp mRNA linear EST 17-OCT-2000
DEFINITION AV731083 HTF Homo sapiens CDNA clone HTFAH01 5', mRNA sequence.
ACCESSION AV731083
VERSION AV731083.1 GI:10840504
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao H., Xu,X.,
Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
Chen,J., Chen,Z. and Han,Z.
TITLE Homo sapiens CDNA HTF clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
INSTITUTION: Chinese National Human Genome Center at Shanghai
ADDRESS: 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
TEL: 86-21-50801919 (ex.45)
FAX: 86-21-50801922
EMAIL: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source 1..736
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTFAH01"
/clone_id="HTF"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); site_1: EcoRI; site_2:
XhoI"

BASE COUNT 193 a 147 c 185 g 211 t

ORIGIN

Query Match 41.5%; Score 208; DB 10; Length 736;
Best Local Similarity 74.9%; Pred. No. 2.9e-55;
Matches 274; Conservative 0; Mismatches 90; Indels 2; Gaps 1;

QY 1 CAAGATCTCAGGATTCATGAGGCTGTGTCTCTATAGCCAGCTGTAGCCCT 60
DB 376 CAAGATCTTACGCTCATGAGGAGGAGTATCTCTATATCCAGTTGTACCAACCC 317
QY 61 TATATCTGCTTTCCCAATATCCAGAGACAGAGTGTATACAGTCTGACCTTCA 120
DB 316 TATATCCCTGCTCTCTCAATATCCAGAGAGAGAGAGTGTATACAGTCTGACCTTCA 257
QY 121 GATGCTCTCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 180
DB 256 GATGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAT 197
QY 181 ACTCAACCAACATCTCAATCTCACTGATATTTTACCCAGAGGTTGAGGATAGT 240
DB 196 CCCAGACACACATCACTCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGAT 137
QY 241 CCCATCTATTTGGCCAGGAGGATGAGCCCAAGCTTGAAGCACTCTCA--TACTGAGCA 298
DB 136 CCTCATCTGTTGGTGAAGGAGCTGAGCCCAAGATCTAGGCTCACTTCAAGTCCAGGCACT 77
QY 299 CTGTGCTTGGTGGTGGATGATTTACTTTTGGCCGCCAATTCAGAAACCTGTGGCAT 358
DB 76 CTGGTCTTCAAGTATGATGATGATTTACTTAATGCTTACCATTTCAAGAAACATGCTCT 17
QY 359 CAAGCC 364
DB 16 CGTGCC 11

RESULT 11 537 bp DNA linear GSS 03-NOV-1998
LOCUS AQ270891/c
DEFINITION HS.2047_A1_H02_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2047 Col=3 Row=O, DNA sequence.
ACCESSION AQ270891
VERSION AQ270891.1 GI:3823431
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 537)
Mahaitsa G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
Hood L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
CONTACT: Mahaitsa G.G., Wallace J.C., Hood L.
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2047 Row: O Column: 3
Class: BAC ends
High quality sequence stop: 537.
Location/Qualifiers
1..537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2047 Col=3 Row=O"
/clone_id="CIT Approved Human Genomic Sperm Library D"
/sex="male"

FEATURES
source

BASE COUNT 125 a 97 c 146 g 161 t 8 others
ORIGIN

Query Match 41.2%; Score 206.2; DB 17; Length 537;
Best Local Similarity 90.0%; Pred. No. 9.2e-55;
Matches 243; Conservative 0; Mismatches 21; Indels 6; Gaps 2;

QY 238 AGTCCCATCTATTTGGCCAGGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 297
DB 536 AGCCCATCTATTTGGCCAGGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 477
QY 298 A--CTTGTCTTGGTGGTGGATGATTTACTTTTGGCCGCCAATTCAGAAACCTGTGC 355
DB 476 ACTTGTCTTGGTGGTGGATGATTTACTTTTGGCCGCCAATTCAGAAACCTGTGC 417
QY 356 CATCAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAA 415
DB 416 CATCAAGCCACCAAGGATGATTTACTTTTGGCCGCCAATTCAGAAACCTGTGC 357
QY 416 CCAGAGGCTCACTCTGCTCAACAGAGT---TACTTGGGCTAAATTTATCCAAAGC 471
DB 356 CCAGAGGCTCACTCTGCTCAACAGAGTAAATATCTTAAATATTCAGAAAGC 297
QY 472 ACCAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 501
DB 296 ACCAGAGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 267

RESULT 12 416 bp DNA linear GSS 14-JUL-1998
LOCUS AQ041889/c
DEFINITION CIT-HSP-2335P3.TR CIT-HSP Homo sapiens genomic clone 2335P3, DNA
sequence.
ACCESSION AQ041889
VERSION AQ041889.1 GI:3311083
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 416)
Adams M.D., Rounsley S.D., Zhao S., Field C.E., Bass S., Linher K.,
Simon M. and Venter J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
CONTACT: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers
1..416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2335P3"
/clone_id="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11, Site_1: HindIII, Site_2:
HindIII"

FEATURES
source

BASE COUNT 119 a 85 c 113 g 99 t
ORIGIN

Seq primer: puc 18 forward
High quality sequence stop: 432.
Location/Qualifiers
1. 434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="HT0513"
/dev_stage="Adult"
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 125 a 92 c 120 g 97 t
ORIGIN

Query Match 40.7%; Score 203.8; DB 10; Length 434;
Best Local Similarity 73.3%; Pred. No. 4.8e-54;
Matches 288; Conservative 0; Mismatches 102; Indels 3; Gaps 2;

Qy 1 CAAGATCTCAGGATATCAATGAGCGCTGTGCTCTATAGCCAGCTGATACCTTACCT 60
Db 413 CAAGATCTTAACTCATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 354
Qy 61 TATATCTGCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 353 G-TACTGCTCTCTCAAAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 295
Qy 121 GATGCTTCTTCTGCACTCCCTGATCCTGATCTCAATCTTGTGCTTGTGAGAT 180
Db 294 GATGCTTCTTCTGATTCACCTGCACTGATCTCCAGTTCTCTTGTGCTTGTGAGAT 235
Qy 181 ACTTAACCCCAATCTCACTCACTGATATTTACCCCAAGGCTTCAAGGATGCT 240
Db 234 CCAACAGACACATCTCCCAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 175
Qy 241 CCCCCTATTTGGCCAGGAGATTAAGCCCAAGCTTGAAGCAATCTCA--TACCTGAGCA 298
Db 174 CCTACCTGTTTGGTCAAGGCTTGGCCCAAGATCTGAGCACTTCTGAACTGAGGCACT 115
Qy 299 CTTGCTCTTGGTATGATGATATTTATTTTGGCCGCAATCAAGAACTTGTGCTAT 358
Db 114 CTGGCTCTTCAAGTATGATGATATTTATTTTGGCTACAGTATGGGAGGCTCGTGCAG 55
Qy 359 CAAGCCACCCAGGCGCTCTTCAATTTCTCGCT 391
Db 54 CAGGCTACTAGATCTCTGAACTTCTAGCT 22

RESULT 15 408 bp DNA linear GSS 02-JUL-1998
A0033042/c 408 bp
LOCUS HS 2226 A1 D12 MF CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=2226 Col=23 Row=G, DNA sequence.
ACCESSION A0033042
VERSION A0033042.1 GI:3285230
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 408)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,S., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNALS Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE Contact: Mahairas GG, Wallace JC, Hood L
COMMENT High Throughput Sequencing Center

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2226 row: G column: 23
Class: BAC ends
High quality sequence stop: 408.
Location/Qualifiers
1. 408
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="Plate=2226 Col=23 Row=G"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 115 a 83 c 105 g 98 t 7 others
ORIGIN

Query Match 40.1%; Score 201; DB 17; Length 408;
Best Local Similarity 69.5%; Pred. No. 3.7e-53;
Matches 282; Conservative 0; Mismatches 122; Indels 2; Gaps 1;

Qy 43 CCAGCTGATCCTAGCCCTTATATCTGCTTCCCAATATCCAGAGAGAGAGAGAGAGAG 102
Db 408 CCAAGTGTACCAACTATATACATGCTCTCTCAATATCCAGAGAGAGAGAGAGAG 349
Qy 103 ACAGTCTGACCTTCAAGATGAGCTTCTTCTGATCCCTGATCCTGATCTCAATTC 162
Db 348 ACTGTTTGAACNTGAAGATGCTTCTTATGATCCATGACATCTGTCTGACAGTTT 289
Qy 163 TTGTTGCTTGAAGATCTTCAACCAACATCTCACTGAGATCTTATTTTCC 222
Db 288 CTTTGTCTTGAAGATCCACAGACCAACGTCNCAATTTATCATGAGAGCTTTGCTT 229
Qy 223 CAAGGTTTCAAGATGATCCCATCTATTTTGGCCAGGATTAAGCCCAAGCTTGAAGCA 282
Db 228 CAAAGGTTTGAAGATGATCCCATCTATCTTGTGATGATATTTGGCNCAGATCTAGGCC 169
Qy 283 TCTCTCA--TACCTGAGCACTTGTCTTGGTATGATATTTTGGCCGCGCCAT 340
Db 168 TTCTCAAGTCCAGGACATCTGTCTCTTCAATATGATATTTTGGCTACAGTT 109
Qy 341 TCAAGAACTTGTGCTATCAAGCCCAAGGCTCTTCAATTTCTGCTACCTGTGGC 400
Db 108 TGAAGCTTTCATGCTCAGCAGGCTATCTAGATCTCTGAACTTCTTGAATCAAGGG 49
Qy 401 TACATGTTTCCAAACCAAGGCTCAACTGTCTCAGCAGAGTTA 446
Db 48 TACAAGCATTTAAACGAGAGGCCAGCTCTGCTTACCAAGTCA 3

Search completed: April 17, 2003, 07:35:31
Job time : 1275.84 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 10:07:23 ; Search time 1700.07 Seconds

(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Perfect score: 500

Sequence: 1 cccctggggcggcctcctt.....tgactccatccctcgtgat 500

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vtc:*
38: em_sy:*
39: em_hugo_hum:*
40: em_hugo_mus:*
41: em_hugo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	500	100.0	10499	6	AX007980	AX007980 Sequence
2	500	100.0	56093	6	AX329572	AX329572 Sequence
3	500	100.0	56093	9	HSAC000064	AC000064 Human BAC
4	500	100.0	149194	9	AC007566	AC007566 Homo sapi
5	435	87.0	711	6	AX007997	AX007997 Sequence
6	410	82.0	783	6	AX000970	AX000970 Sequence
7	383.8	76.8	711	6	AX007998	AX007998 Sequence
8	360	72.0	137492	2	AL161721	AL161721 Homo sapi
9	359.8	72.0	143590	2	HS83613	AL161721 Homo sapi
10	358.2	71.6	119406	2	AC103595	AL161721 Homo sapi
11	357.8	71.6	110000	2	AL353584_0	AL353584 Homo sapi
12	357.8	71.6	169029	9	CNS06C7R	AL353584 Homo sapi
13	357.8	71.6	170746	9	AL392173	AL392173 Human chr
14	357.4	71.5	163803	9	AC093531	AC093531 Homo sapi
15	357.4	71.5	191863	9	AC010888	AC010888 Homo sapi
16	356.8	71.4	180635	9	AL360169	AL360169 Homo sapi
17	356.6	71.3	174019	9	AP001538	AP001538 Homo sapi
18	354.2	70.8	340000	9	AP001674	AP001674 Homo sapi
19	354.2	70.8	111079	9	HS419C19	AL035407 Human DNA
20	353.4	70.7	76169	9	AC003014	AC003014 Human DNA
21	351.8	70.4	161049	2	AL591842	AL591842 Homo sapi
22	351	70.2	42216	9	AC092899	AC092899 Homo sapi
23	351	70.2	170412	2	AC026019	AC026019 Homo sapi
24	349.6	69.9	83412	9	AC092843	AC092843 Homo sapi
25	349.2	69.8	165059	9	AC106856	AC106856 Homo sapi
26	348.2	69.6	839	9	AF127226	AF127226 Homo sapi
27	348.2	69.6	1329	6	AX001030	AX001030 Sequence
28	347.6	69.5	177212	2	AC025733	AC025733 Homo sapi
29	347.6	69.5	181753	9	AC124075	AC124075 Homo sapi
30	347.2	69.4	119481	9	AL136234	AL136234 Human DNA
31	347.2	69.4	152996	9	AC013719	AC013719 Homo sapi
32	346.4	69.3	179585	2	AC024977	AC024977 Homo sapi
33	346.4	69.3	185671	9	AC084193	AC084193 Homo sapi
34	346.2	69.2	156177	9	AC093896	AC093896 Homo sapi
35	345	69.0	37764	9	HS0221F2	275746 Human DNA s
36	345	69.0	143620	9	AC117425	AC117425 Homo sapi
37	344.8	69.0	157711	9	AC055738	AC055738 Homo sapi
38	344.8	69.0	181850	9	AC018680	AC018680 Homo sapi
39	344.8	69.0	207945	2	CNS01DSS	AL121841 Human chr
40	344.8	69.0	318442	2	AC021652	AC021652 Homo sapi
41	344.6	68.9	140756	9	AL139038	AL139038 Human DNA
42	344.6	68.9	183499	2	AL607153	AL607153 Homo sapi
43	344.4	68.9	88328	9	AL357874	AL357874 Human DNA
44	343	68.6	187321	2	AC092510	AC092510 Papio cyn
45	341.6	68.3	102200	9	HS30M3	AL031775 Human DNA

ALIGNMENTS

RESULT 1
AX007980
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 10499)
AUTHORS
TITLE
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses

JOURNAL

Patent: WO 9967395-A 3 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERLIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)

FEATURES

Location/Qualifiers
1..10499
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 2495 t
ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 10499;

Best Local Similarity 100.0%; Pred. No. 8.5e-155;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCTGTGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCTGGAGATACAGCAATTATC 60
DB 1 CCTGTGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCTGGAGATACAGCAATTATC 60
QY 61 TTGCAACTGAGAGACAGGACTAGCTGATTTCTTAG;CCGACTAAGAAATCCCTAAGCCTA 120
DB 61 TTGCAACTGAGAGACAGGACTAGCTGATTTCTTAG;CCGACTAAGAAATCCCTAAGCCTA 120
QY 121 GCTGGGAAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGGCACTTAGCTCACACTGA 180
DB 121 GCTGGGAAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGGCACTTAGCTCACACTGA 180
QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGCAGAGATTAAGAAATAGCCA 240
DB 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGCAGAGATTAAGAAATAGCCA 240
QY 241 ATCATCTATTGCTGAGAGACAGAGAGGAGCAACATCGGATATATAAACCCAGGCAT 300
DB 241 ATCATCTATTGCTGAGAGACAGAGAGGAGCAACATCGGATATATAAACCCAGGCAT 300
QY 301 TCGAGCTGGCAACAGAGCCCCCTTTGGGTCCCTTC;CTTTGATGGAGCTGTTTCA 360
DB 301 TCGAGCTGGCAACAGAGCCCCCTTTGGGTCCCTTC;CTTTGATGGAGCTGTTTCA 360
QY 361 TGCATTTTCACTCTATTAAATCTTGCAACTGCACTCT;CTGCTCATGTTTCTTACGGCT 420
DB 361 TGCATTTTCACTCTATTAAATCTTGCAACTGCACTCT;CTGCTCATGTTTCTTACGGCT 420
QY 421 CGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTT GCCACACCGCAGACTGCGGC 480
DB 421 CGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTT GCCACACCGCAGACTGCGGC 480
QY 481 TGACTCCCATCCCTCTGGAT 500
DB 481 TGACTCCCATCCCTCTGGAT 500
```

RESULT 2

AX329572 56093 bp DNA linear PAT 09-JAN-2002

LOCUS AX329572 Sequence 81 from Patent WO0194629.

DEFINITION AX329572

ACCESSION AX329572

VERSION AX329572.1 GI:18102550

KEYWORDS

SOURCE

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

TITLE

Young, P.E., Augustus, M., Carter, K.C., Ehmer, R., Endress, G.,
Horrikan, S., Soper, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets

JOURNAL

Patent: WO 0194629-A 81 13-DEC-2001;

FEATURES

SOURCE

1..56093
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 56093;

Best Local Similarity 100.0%; Pred. No. 9.4e-155;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCTGTGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCTGGAGATACAGCAATTATC 60
DB 28001 CCTGTGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCTGGAGATACAGCAATTATC 28060
QY 61 TTGCAACTGAGAGACAGGACTAGCTGATTTCTTAGGCCGACTAAGAAATCCCTAAGCCTA 120
DB 28061 TTGCAACTGAGAGACAGGACTAGCTGATTTCTTAGGCCGACTAAGAAATCCCTAAGCCTA 120
QY 121 GCTGGGAAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGGCACTTAGCTCACACTGA 180
DB 28121 GCTGGGAAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGGCACTTAGCTCACACTGA 180
QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGCAGAGATTAAGAAATAGCCA 240
DB 28181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGCAGAGATTAAGAAATAGCCA 28240
QY 241 ATCATCTATTGCTGAGAGACAGAGAGGAGCAACATCGGATATATAAACCCAGGCAT 300
DB 28241 ATCATCTATTGCTGAGAGACAGAGAGGAGCAACATCGGATATATAAACCCAGGCAT 28300
QY 301 TCGAGCTGGCAACAGAGCCCCCTTTGGGTCCCTTC;CTTTGATGGAGCTGTTTCA 360
DB 28301 TCGAGCTGGCAACAGAGCCCCCTTTGGGTCCCTTC;CTTTGATGGAGCTGTTTCA 28360
QY 361 TGCATTTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGCTCATGTTTCTTACGGCT 420
DB 28361 TGCATTTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGCTCATGTTTCTTACGGCT 28420
QY 421 CGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTT GCCACACCGCAGACTGCGGC 480
DB 28421 CGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTT GCCACACCGCAGACTGCGGC 28480
QY 481 TGACTCCCATCCCTCTGGAT 500
DB 28481 TGACTCCCATCCCTCTGGAT 28500
```

RESULT 3

HSAC000064 56093 bp DNA linear PRI 13-NOV-1996

LOCUS HSAC000064 Human BAC clone RG083M05 from 7q21-7q22, complete sequence.

DEFINITION AC000064

ACCESSION AC000064

VERSION AC000064.1 GI:1669369

KEYWORDS

SOURCE

ORGANISM

Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

TITLE

The sequence of H. sapiens BAC clone RG083M05
Unpublished (1996)

JOURNAL

REFERENCE

AUTHORS

TITLE

Waterson, R.
Direct Submission
Submitted (13-NOV-1996)

JOURNAL

Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA

e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
sections once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The
library contains cloned DNA from a human male fibroblast cell line
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBELO

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this
clone is at base position 1 of H_RG083M05; actual end is at 56093
of H_RG083M05

FEATURES

source

This clone contains SRS SMS51725.

Location/Qualifiers

1. 56093

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone="H_RG083M05"

/clone_1lb="CITB-978SK-B"

/complement(838..1131)

/rpt_family="ALU"

<1360..16971

/gene="WUGSC:H_RG083M05.1"

join(<1360..1503,4181..4370,4587..4774,6422..6556,

9483..9547,11631..11773,11864..12021,13131..13296,

14885..14988,16349..16546,16837..16971)

/gene="WUGSC:H_RG083M05.1"

/note="Atpase; strong similarity to peroxisome

biosynthesis protein PAB1 (PID:g117019); coded for by

human cDNA C04279 (NID:g1467550)"

/protein_start=1

/protein_id="AAB46346.1"

/db_xref="GI:1669371"

/translation="KLENTIOKTELEVAFSEAVMMOPSVLLDDLDLIAGLPAVBEHEH

SPAPBORCEICNVIKKIDCDGINKFTDLDLGVAKETGFAVARDFTVDRALHRL

SRSISITREKLVTTIDDPQKALRGFPASLRSSNLUKPRDLGDKIGGHEVQIILMD

T10LPKAVCLKKEKYPELEFANLP1RQRTGILYDPGEGTKTLGAVIARSRNFTSV

KGBELLSKYIGABEQAVRDIPIRQAQAKPCILFDFEESJAPRRDHNTGTVRVNQ

LTLQDLGVEGLQGVYVLAATSRPDLIDPALLRGRIDKCYCPPOVITISYESTKO

OMHSFLVSRLEILNVLSDELPLADVDLQHVASVDSFGADLKALLVVAOLEALHG

MLSKMSRILPDSKFMVRLYFGSSYSESLGNGTSDDLSSOCLSA PSSWTODLPVP

GKQULFSQPVYLTASQSCQELTQGEORDDLRADISTIKRYSQSGEDSMNQPGPI

KTRLAISQSHLMALGHTRPSISEDWKNFAEL"

/complement(4948..5130)

/rpt_family="ALU"

/complement(6581..7133)

/rpt_family="U1"

/complement(7767..8037)

/rpt_family="ALU"

/complement(8186..8472)

/rpt_family="ALU"

8473..8625

/gene="WUGSC:H_RG083M05.1"

/note="match to human 3' EST H75782 (NID:g1049794), bases

287-444"

8841..9161

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H75921 (NID:g1050050), bases

21-348"

9481..9547

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST N22627 (NID:g1130501), bases

276-343"

repeat_region

complement(12612..12907)

/rpt_family="ALU"

13670..13793

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

143-266"

13794..13877

/rpt_family="ALU"

13878..13906

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

30-58"

13907..14104

/rpt_family="ALU"

complement(14110..14137)

/rpt_family="U1"

complement(15618..15907)

/rpt_family="ALU"

17227..17522

/rpt_family="ALU"

18667..19235

/note="match to human fetal brain 5' EST D61494

(NID:g970409), bases 1-255, and to human 3' EST R07476

(NID:g759399)"

19550..19670

/rpt_family="ALU"

21507..37303

/note="similarity to various SS-RNA virus polypeptides;

pseudogene; region of matches and close matches to

multiple human ESTs, see R68740 (NID:g942257)"

37316..37489

/note="Grail prediction, score = 80"

/evidence=not experimental

complement(138538..139224)

/rpt_family="ALU"

39225..39707

/note="match to multiple human ESTs, see N30113

(NID:g1148633)"

39800..40085

/rpt_family="ALU"

complement(40247..40538)

/rpt_family="ALU"

complement(40632..40924)

/rpt_family="ALU"

complement(42283..42891)

/rpt_family="ALU"

complement(45474..45613)

/rpt_family="ALU"

complement(45614..45737)

/note="match to human 3' EST H48898 (NID:g988738), bases

129-333"

complement(46107..47026)

/note="match to multiple human ESTs, see N81064

(NID:g1243765), H48897 (NID:g988737), and M78831

(NID:g273146)"

complement(47027..47318)

/rpt_family="ALU"

complement(47365..47782)

/note="match to multiple human ESTs, see W37495

(NID:g1319089)"

47898..48115

/note="match to human 5' EST H62306 (NID:g1015138), bases

93-368"

complement(48116..48405)

/rpt_family="ALU"

complement(48406..48584)

/note="match to human 3' EST N29952 (NID:g1148472), bases

250-455, and 5' EST R12730 (NID:g765806)"

complement(48787..49405)

/rpt_family="ALU"

complement(49406..49534)

/note="match to human 3' EST R65794 (NID:g838432), bases

repeat_region

complement(12612..12907)

/rpt_family="ALU"

13670..13793

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

143-266"

13794..13877

/rpt_family="ALU"

13878..13906

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

30-58"

13907..14104

/rpt_family="ALU"

complement(14110..14137)

/rpt_family="U1"

complement(15618..15907)

/rpt_family="ALU"

17227..17522

/rpt_family="ALU"

18667..19235

/note="match to human fetal brain 5' EST D61494

(NID:g970409), bases 1-255, and to human 3' EST R07476

(NID:g759399)"

19550..19670

/rpt_family="ALU"

21507..37303

/note="similarity to various SS-RNA virus polypeptides;

pseudogene; region of matches and close matches to

multiple human ESTs, see R68740 (NID:g942257)"

37316..37489

/note="Grail prediction, score = 80"

/evidence=not experimental

```

repeat_region      complement(49638..49672)
                    /rpl_family="ALU"
misc_feature        complement(49674..49890)
                    /note="match to human 3' EST N29952 (NID:g1148472) and 5'
                    EST N29938 (NID:g1148450), sequences are from opposite
                    ends of the same clone"
gene                complement(49698..51806)
                    /gene="WUGSC:H.RG083M05.2"
                    /complement(join(49698..49888,51575..51806))
                    /gene="WUGSC:H.RG083M05.2"
                    /note="coded for by human CDNA W37389 (NID:g119205),
                    R65891 (NID:g838529), R65794 (NID:g838432) and R65794
                    (NID:g838432)"
CDS                 /protein_id="AAB46345.1"
                    /codon_start=1
                    /db_xref="GI:169370"
                    /translation="MFYFYRQCGTTFPCPGVYVYQIDVYVIDEDQKPYVQIRGF
                    IQDYCKSAALTLIPLTSPRD)PDPASTITGPEDLPKMYELFVCHABSEYTK
                    SRSPFPVTPRPKGYIWHVGP:PAITIKESVANHL"
exon                complement(51576..51758)
                    /gene="WUGSC:H.RG083M05.2"
                    /note="Grail prediction, score = 86"
repeat_region       /evidence=not_experimental
                    complement(52052..52329)
                    /rpl_family="L1"
                    /note="match to human EST M79192 (NID:g273505) base 2-289"
misc_feature        55557..55843
Query Match         100.0%; Score 500; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. No. 9,4e-155;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCTGGGGGGGGCTTCCTTCTGGAGTGGGCAAAAGCCTGGAGATACAGCAATATC 60
Db 28001 CCTGGGGGGGGCTTCCTTCTGGAGTGGGCAAAAGCCTGGAGATACAGCAATATC 28060
Qy 61 TTGCAACTGAGAGACAGAGACTAGTGGATTCTTCTAGCCGACTAAGAAATCCCTAAGCCTA 120
Db 28061 TTGCAACTGAGAGACAGAGACTAGTGGATTCTTCTAGCCGACTAAGAAATCCCTAAGCCTA 28120
Qy 121 GCTGGGAAGGTGACCAACGCTTCAACCGGGGTTTGCACTTAGCTACACCTTGA 180
Db 28121 GCTGGGAAGGTGACCAACGCTTCAACCGGGGTTTGCACTTAGCTACACCTTGA 28180
Qy 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAACAGAGAGTAAAGAAATAGCCA 240
Db 28181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAACAGAGAGTAAAGAAATAGCCA 28240
Qy 241 ATCATCTATTGCTTGAAGACAGACAGAGAGGAGCAACATCGGGATATTAAACCCAGGCAT 300
Db 28241 ATCATCTATTGCTTGAAGACAGACAGAGAGGAGCAACATCGGGATATTAAACCCAGGCAT 28300
Qy 301 TCGAGCTGGCAACAGACAGCCCCCTTTGGGTCCCTTCTTGATTTGGAGTGTCTTTCA 360
Db 28301 TCGAGCTGGCAACAGACAGCCCCCTTTGGGTCCCTTCTTGATTTGGAGTGTCTTTCA 28360
Qy 361 TGCATTTCACCTCTAATAATCTTGCAACTGCACTCTTGTGTCATGTCTTTACGGCT 420
Db 28361 TGCATTTCACCTCTAATAATCTTGCAACTGCACTCTTGTGTCATGTCTTTACGGCT 28420
Qy 421 CGAGCTGAGCTTTTCTCCACGCTGACCACTGCTGTTTTCACCAACCGAGACTGCGCG 480
Db 28421 CGAGCTGAGCTTTTCTCCACGCTGACCACTGCTGTTTTCACCAACCGAGACTGCGCG 28480
Qy 481 TGACTCCCATCCCTCTGGAT 500
Db 28481 TGACTCCCATCCCTCTGGAT 28500

```

```

RESULT 4
AC007566/c AC007566 149194 bp DNW linear PRI 01-MAR-2002
LOCUS      Homo sapiens BAC clone CTB-10G5 from 'g21-7q22, complete sequence.
DEFINITION

```

```

ACCESSION AC007566
VERSION AC007566.2 GI:1181861
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
          1 (bases 1 to 149194)
          Sulston, J.E. and Waterston, R.
          Toward a complete human genome sequence
          Genome Res. 8 (11), 1097-1108 (1998)
          99063792
          9847074
MEDLINE 2 (bases 1 to 149194)
PUBMED 2.
REFERENCE Du, Z.
          The sequence of Homo sapiens BAC clone CTB-10G5
          Unpublished (2001)
          3 (bases 1 to 149194)
REFERENCE Waterston, R.H.
          Direct Submission
          Submitted (15-MAY-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          4 (bases 1 to 149194)
REFERENCE Waterston, R.
          Direct Submission
          Submitted (02-OCT-2000) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          5 (bases 1 to 149194)
REFERENCE Waterston, R.H.
          Direct Submission
          Submitted (16-NOV-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          6 (bases 1 to 149194)
REFERENCE Waterston, R.H.
          Direct Submission
          Submitted (03-JAN-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          7 (bases 1 to 149194)
REFERENCE Waterston, R.H.
          Direct Submission
          Submitted (06-FEB-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          8 (bases 1 to 149194)
REFERENCE Waterston, R.
          Direct Submission
          Submitted (01-MAR-2002) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          On Nov 16, 2000 this sequence version replaced gi:4835815.
COMMENT ----- Genome Center
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc
          Contact: sapiens@watson.wustl.edu
          ----- Summary Statistics
          -----
          Center project name: H_RG010G05

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-9768SK-B. The library contains cloned DNA from the male fibroblast cell line 9768SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 155 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

source

```

1.149194
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="7"
  /map="7q21-7q22"
  /clone_1b="CTB-10G5"
  /clone_1b="CTB-9768SK-B"
  1.1634
    /rpt_family="L1"
      repeat_region
        misc_feature
          2248..2388
            /note="match to EST BG752883 (NID:g14063536) "
        misc_feature
          2248..2387
            /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
        misc_feature
          2248..2387
            /note="match to EST AM579261 (NID:g7254310) "
        misc_feature
          2248..2387
            /note="match to EST BG766882 (NID:g14077535) "
        misc_feature
          2248..2387
            /note="match to EST B1160365 (NID:g14620366) "
        misc_feature
          2248..2374
            /note="similar to Homo sapiens EST BF758865 (NID:g12106765) "
        misc_feature
          2248..2287
            /note="match to EST AV686676 (NID:g10288539) "
        misc_feature
          2253..2387
            /note="match to EST AU123510 (NID:g10948226) "
        misc_feature
          2344..2387
            /note="match to EST AV686676 (NID:g10288539) "
        repeat_region
          2696..3066
            /rpt_family="Mair"
              repeat_region
                misc_feature
                  3108..3392
                    /rpt_family="Alu"
                      misc_feature
                        3540..3628
                          /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
                        3540..3628
                          /note="similar to Mus musculus EST BB253526 (NID:g8946272) "
                      misc_feature
                        3542..3632
                          /note="match to EST AV686676 (NID:g10288539) "
                      misc_feature
                        3542..3628
                          /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
                      misc_feature
                        3542..3628
                          /note="match to EST AU123510 (NID:g10948226) "
                      misc_feature
                        3542..3628
                          /note="match to EST BG766882 (NID:g14077535) "

```

```

misc_feature
  3542..3628
    /note="match to EST B1160365 (NID:g14620366) "
misc_feature
  3542..3609
    /note="match to EST AM579261 (NID:g7254310) "
misc_feature
  3562..3628
    /note="match to EST BE272564 (NID:g9146913) "
misc_feature
  3598..3628
    /note="match to EST BE299708 (NID:g9183456) "
misc_feature
  3714..3785
    /note="match to EST BG260659 (NID:g12770475) "
misc_feature
  3717..3785
    /note="similar to Mus musculus EST BE994936 (NID:g10678674) "
misc_feature
  3719..3785
    /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
  3719..3785
    /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
misc_feature
  3719..3785
    /note="match to EST AU123510 (NID:g10948226) "
misc_feature
  3719..3785
    /note="match to EST BE272564 (NID:g9146913) "
misc_feature
  3719..3785
    /note="match to EST BE299708 (NID:g9183456) "
misc_feature
  3719..3785
    /note="match to EST BG766882 (NID:g14077535) "
misc_feature
  3719..3785
    /note="match to EST B1160365 (NID:g14620366) "
misc_feature
  3719..3785
    /note="similar to Mus musculus EST BB253526 (NID:g8946272) "
misc_feature
  3860..3862
    /note="match to EST AV686676 (NID:g10288539) "
repeat_region
  4730..4827
    /rpt_family="L2"
      repeat_region
        misc_feature
          4834..4925
            /rpt_family="L1"
              misc_feature
                5715..5860
                  /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
              misc_feature
                5715..5860
                  /note="match to EST AU123510 (NID:g10948226) "
              misc_feature
                5715..5860
                  /note="match to EST BE272564 (NID:g9146913) "
              misc_feature
                5715..5860
                  /note="match to EST BE299708 (NID:g9183456) "
              misc_feature
                5715..5860
                  /note="match to EST BG260659 (NID:g12770475) "
              misc_feature
                5715..5860
                  /note="match to EST BG766882 (NID:g14077535) "
              misc_feature
                5715..5860
                  /note="match to EST B1160365 (NID:g14620366) "
              misc_feature
                5715..5860
                  /note="similar to Mus musculus EST BE994936

```

Query Match 100.0%; Score 500; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 1e-154;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CCTGGGGGCGGCTTCCTTCTGGATGAGGGCAAAACGCTTGAAGTATAGCAATTATC 60
DB 94003 CCTGGGGGCGGCTTCCTTCTGGATGAGGGCAAAACGCTTGAAGTATAGCAATTATC 93944
QY 61 TTGCAACTGAGAGACAGAGTACGATTTCTTCTAGGCGAGCTAAGATCCCTAAGCCTA 120
DB 93943 TTGCAACTGAGAGACAGAGTACGATTTCTTCTAGGCGAGCTAAGATCCCTAAGCCTA 93884
QY 121 GCTGGAGAGGTGACCACTGACACCTTTAAACAGGGGCTTGCACTTAGCTCACCTGA 180
DB 93883 GCTGGAGAGGTGACCACTGACACCTTTAAACAGGGGCTTGCACTTAGCTCACCTGA 93824
QY 181 CCAATCAGAGAGCTCACTTAATATGCTTAATAGGCAAGACAGAGGTAAAGAAATAGCCA 240
DB 93823 CCAATCAGAGAGCTCACTTAATATGCTTAATAGGCAAGACAGAGGTAAAGAAATAGCCA 93764

```

QY 241 ATCATATTGCTGAGAGACAGAGAGAGACACAAATCGGGATATATAAACCAGGCAT 300
Db 93763 ATCATATTGCTGAGAGACAGAGAGAGAGACACAAATCGGGATATATAAACCAGGCAT 93704
QY 301 TCGAGCTGGCAACAGACACCCCTTTGGGTCCTTCCCTTTGTAATGGAGCTGTTTCA 360
Db 93703 TCGAGCTGGCAACAGACACCCCTTTGGGTCCTTCCCTTTGTAATGGAGCTGTTTCA 93644
QY 361 TCGATTTCACCTATTAAATCTTGAACCTGCACTTTCTGGTCCATGTTTCTTAAGGCT 420
Db 93643 TCGATTTCACCTATTAAATCTTGAACCTGCACTTTCTGGTCCATGTTTCTTAAGGCT 93584
QY 421 CGAGCTGAGCTTTTGGCTCAGCCGTCACCACTGCTGTTGGCAACACCCGAGACTGCGGC 480
Db 93583 CGAGCTGAGCTTTTGGCTCAGCCGTCACCACTGCTGTTGGCAACACCCGAGACTGCGGC 93524
QY 481 TGACTCCCATCCCTCTGGAT 500
Db 93523 TGACTCCCATCCCTCTGGAT 93504

RESULT 5
AX007997
LOCUS AX007997 711 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 20 from Patent WO9667395.
ACCESSION AX007997
VERSION AX007997.1 GI:995694
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 711)
TITLE Perin, J.P., Rieger, F. and Alliel, P.M.
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
JOURNAL Patent: WO 9667395-A 20 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES
LOCATION/Qualifiers
SOURCE 1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 170 a 204 c 162 g 175 t
ORIGIN

Query Match 87.0%; Score 435; DI: 6; Length 711;
Best Local Similarity 100.0%; Pred. No. 3,7e-133; Indels 0; Gaps 0;
Matches 435; Conservative 0; Mismatches 0;

QY 66 ACTGAGAGACAGACTAGCTGATTTCTAGGCGCACTAAGAATCCCTAAGCTAGCTGG 125
Db 1 ACTGAGAGACAGACTAGCTGATTTCTAGGCGCACTAAGAATCCCTAAGCTAGCTGG 60
QY 126 GAAGTGACCACTGCACTTTAAACACGGGCTTGCACTTAACTGACACCTGACCAAT 185
Db 61 GAAGTGACCACTGCACTTTAAACACGGGCTTGCACTTAACTGACACCTGACCAAT 120
QY 186 CAGAGAGCTCAATAAATCTAATTAAGGAGAGACAGAGTAAGAAATAGCAATCAT 245
Db 121 CAGAGAGCTCAATAAATCTAATTAAGGAGAGACAGAGTAAGAAATAGCAATCAT 180
QY 246 CTATTGCTGAGAGACAGAGAGAGAGCAACATCGCATATAAACCAGGACTTCGAG 305
Db 181 CTATTGCTGAGAGACAGAGAGAGAGCAACATCGCATATAAACCAGGACTTCGAG 240
QY 306 CTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTTGATAGGAGCTGTTTCACTGA 365
Db 241 CTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTTGATAGGAGCTGTTTCACTGA 300
QY 366 TTTCACTCTAATTAATCTTGAACCTGCACTTTCTGGTCCATGTTTCTTAAGGCTGAGC 425
Db 301 TTTCACTCTAATTAATCTTGAACCTGCACTTTCTGGTCCATGTTTCTTAAGGCTGAGC 360

QY 426 TGAGCTTTTGTCAACCCGTCACCACTGCTGTTTGGCAACACCCGAGACTGCGCTACT 485
Db 361 TGAGCTTTTGTCAACCCGTCACCACTGCTGTTTGGCAACACCCGAGACTGCGCTACT 420
QY 486 CCCATCCCTCTGGAT 500
Db 421 CCCATCCCTCTGGAT 435

RESULT 6
AX000970
LOCUS AX000970 783 bp DNA linear PAT 10-MAR-2000
DEFINITION Sequence 15 from Patent WO902696.
ACCESSION AX000970
VERSION AX000970.1 GI:7241212
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 783)
AUTHORS Beseme, F. and Blond, J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 9902696-A 15 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
LOCATION/Qualifiers
SOURCE 1..783
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 173 a 213 c 166 g 180 t 51 others
ORIGIN

Query Match 82.0%; Score 410; DI: 6; Length 783;
Best Local Similarity 93.8%; Pred. No. 6.8e-125;
Matches 408; Conservative 24; Mismatches 1; Indels 2; Gaps 1;

QY 68 TGAGAGACAGACTAGCTGATTTCTAGGCGCACTAAGAATCCCTAAGCTAGCTGG 127
Db 1 TGAGAGACAGACTAGCTGATTTCTAGGCGCACTAAGAATCCCTAAGCTAGCTGG 60
QY 128 AGGAGACCACTGCACTTTAAACACGGGCTTGCACTTAACTGACACCTGACCAATCA 187
Db 61 AGGAGACCACTGCACTTTAAACACGGGCTTGCACTTAACTGACACCTGACCAATCA 120
QY 188 GAGAGCTCAATAAATCTAATTAAGGAGAGACAGAGTAAGAAATAGCAATCATCT 247
Db 121 GAGAGCTCAATAAATCTAATTAAGGAGAGACAGAGTAAGAAATAGCAATCATCT 180
QY 248 ATTGCTGAGAGACAGAGAGAGAGCAACATCGGATATATAAACCAGGCACTGAGCT 307
Db 181 ATTGCTGAGAGACAGAGAGAGAGCAACATCGGATATATAAACCAGGCACTGAGCT 240
QY 308 GGCAACAGACACCCCTTTGGGTCCCTTCCCTTTGATAGGAG--CTGTTTCACTGCTA 365
Db 241 GGCAACAGACACCCCTTTGGGTCCCTTCCCTTTGATAGGAGCTGTTTCACTGCTA 300
QY 366 TTTCACTCTAATTAATCTTGAACCTGCACTTTCTGGTCCATGTTTCTTAAGGCTGAGC 425
Db 301 TTTCACTCTAATTAATCTTGAACCTGCACTTTCTGGTCCATGTTTCTTAAGGCTGAGC 360
QY 426 TGAGCTTTTGTCAACCCGTCACCACTGCTGTTTGGCAACACCCGAGACTGCGCTACT 485
Db 361 TGAGCTTTTGTCAACCCGTCACCACTGCTGTTTGGCAACACCCGAGACTGCGCTACT 420
QY 486 CCCATCCCTCTGGAT 500
Db 421 CCCATCCCTCTGGAT 435

RESULT 7
AX007998
LOCUS AX007998 711 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 21 from Patent WO967395.
ACCESSION AX007998
VERSION AX007998.1 GI:9995695
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 711)
AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
Patent: WO 96/7395-A 21 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES
source 1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 171 a 209 c 160 g 171 t
ORIGIN
Query Match 76.8%; Score 383.8; DB 6; Length 711;
Best Local Similarity 95.4%; Pred.No.3.6e-116;
Matches 417; Conservative 0; Mismatches 17; Indels 3; Gaps 2;
QY 66 ACTGAGAGACAGAGCTGATGATTTCTTCTAGCCGACCTAAGATCCCTAAGCTAGCTGG 125
DB 1 ACTGAGAGACAGAGCTGATGATTTCTTCTAGCCGACCTAAGATCCCTAAGCTAGCTGG 60
QY 126 GAAAGTACCAACGTCACCTTTAAACAGGGGGCTTGGCACTTACCTCAACCTGACCAAT 185
DB 61 GAAAGTACCAACGTCACCTTTAAACAGGGGGCTTGGCACTTACCTCAACCTGACCAAT 120
QY 186 CAGAGAGCTACTTAATGCTAATTAGCAAAAGCAGAGGTAAAGAAATAGCCAAATCAT 245
DB 121 CAGAGAGCTACTTAATGCTAATTAGCAAAAGCAGAGGTAAAGAAATAGCCAAATCAT 180
QY 246 CTATGCTGAGACAGACAGAGAGGAGCAACATCGGATATTAACCCAGGACTTGGAG 305
DB 181 CTATGCTGAGACAGACAGAGAGGAGCAACATCGGATATTAACCCAGGACTTGGAG 240
QY 306 CTGCGCAACAGACAGCCCTTTGGGTCCTTCCCTTTGATGGAG--CTGTTTTCATGC 363
DB 241 CCGGCAACGCGCA-ACCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTTCATGC 299
QY 364 TATTTACTTATTAATCTTTCACATGCACTCTTCTGTCATGTTTCTTACGGCTCGA 423
DB 300 TATTTACTTATTAATCTTTCACATGCACTCTTCTGTCATGTTTCTTACGGCTCGA 359
QY 424 GCTGAGCTTTGCTCAACCGTCACCACTGCTGTTGACACCGCAGACCTGCGCTGA 483
DB 360 GCTGAGCTTTGCTGCGCATTCACCACTGCTGTTGCGCCACCGCAGACCGCGCTGA 419
QY 484 CTCCCATCCCTCGAT 500
DB 420 CTCCCATCCCTCGAT 436
RESULT 8
AL161721/c 137492 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 6 clone RPI-301F24 map q25.2-26, ***
DEFINITION SEQUENCING IN PROGRESS ***, 6 unordered pieces.
ACCESSION AL161721 GI:12330750
VERSION AL161721.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 137492)
REFERENCE 1
AUTHORS Sims,S.

TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Jan 22, 2001 this sequence version replaced gi:9212839.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj301F24
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 13573 bases at least Q40
Consensus quality: 13524 bases at least Q30
Consensus quality: 13616 bases at least Q20
Insert size: 136992; sum-of-contigs
Insert size: 152637; 0.5% error; agarose-fp
Quality coverage: 3.96x in Q20 bases; sum-of-contigs Quality
Coverage: 3.56x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12245: contig of 12245 bp in length
* 12246 12345: gap of 100 bp
* 12346 20428: contig of 8083 bp in length
* 20429 20528: gap of 100 bp
* 20529 85104: contig of 64576 bp in length
* 85105 85204: gap of 100 bp
* 85205 96812: contig of 11608 bp in length
* 96813 96912: gap of 100 bp
* 96913 112622: contig of 15710 bp in length
* 112623 112722: gap of 100 bp
* 112723 137492: contig of 24770 bp in length.
* Location/Qualifiers
1..137492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q25.2-26"
/clone_1fb="RPI-1"
1..12245
/note="assembly_fragment:00933
fragment_chain:1
clone_end:17
vector_side:left"
12346..20428
/note="assembly_fragment:01071
fragment_chain:1"
20529..85104
/note="assembly_fragment:00093
fragment_chain:1"
85205..96812
/note="assembly_fragment:00954
fragment_chain:1"
96913..112622
/note="assembly_fragment:01341"
112723..137492
/note="assembly_fragment:00790
clone_end:866
vector_side:right"
BASE COUNT 37669 a 28026 c 30367 g 40929 t 501 others
ORIGIN

Query Match 72.0%; Score 360; DB 2; Length 137492;
 Best Local Similarity 89.6%; Pred. No. 4,1e-108; Indels 1; Gaps 1;
 Matches 399; Conservative 0; Mismatches 45;

Qy 57 TATCTTGCACTGAGAGACAGACTAGCTGATTTCTTCCAGCCGACTAAGATCCCTAAG 116
 |||||
 Db 128053 TTTTATAGTTCTGAGAGACAGACTAGCTGATTTCTTCCAGCCGACTAAGATCCCTAAG 127994

Qy 117 CCTTGGCTGGAGAGTGAACAGCTCCACTTTAAACA-3GGGGCTTGCACTTAAGCTCACAC 176
 |||||
 Db 127993 CCTTGGCTGGAGAGTGAACAGCTCCACTTTAAACA-3GGGGCTTGCACTTAAGCTCACAC 127934

Qy 177 CTGACCAATCAGAGAGCTCAGTAATGCTAATTAG;CAAGAAGAGAGTAAAGAAATA 236
 |||||
 Db 127933 CCAACCAATCAGAGAGCTCAGTAATGCTAATTAG;CAAGAAAGAGAGTAAAGAAATA 127874

Qy 237 GCCAATCATCTATTGCTTGAAGACAGACAGAGAGG;CAACATCGGATATTAACCCAG 296
 |||||
 Db 127873 GCCAATCATCTATTGCTTGAAGACAGACAGAGAGG;CAAGATCGGATATTAACCCAG 127814

Qy 297 GCATTGAGCTGGCAACAGACCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTT 356
 |||||
 Db 127813 GCATTGAGCTGGCAACAGCA-ACCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTT 127755

Qy 357 TTCAATGCTATTCTACTATTAAATCTTGCAACTGCTCTTGGTCCATGTTCTTAC 416
 |||||
 Db 127754 TTCAATGCTATTCTACTATTAAATCTTGCAACTGCACTCTTGGTCCATGTTCTTAC 127695

Qy 417 GGCTGAGCTGAGCTTTTGTCTACCGCTCCACCACTGCTGTTTGCACACCGCAGACCTG 476
 |||||
 Db 127694 GGCTGAGCTGAGCTTTTGTCTGCTGCTGCTCCACCACTGCTGTTTGCACACCGCAGACCTG 127635

Qy 477 CCGCTGATCCCATCTCTCTGAT 500
 |||||
 Db 127634 CCGCTGATCCCATCTCTCTGAT 127611

RESULT 9
 HS836J3/c 143590 bp LNA linear PRI 02-MAR-2000
 LOCUS Human DNA sequence from clone RP5-836J3 on chromosome 1p22.1-31.1
 DEFINITION Contains STSs, GSSs and a putative CpG island, complete sequence.
 ACCESSION AL035706
 VERSION AL035706.10 GI:5777502
 KEYWORDS HTG; CpG island.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 143590)
 Bagguley, C.
 Direct Submission
 Submitted (29-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequest@sanger.ac.uk
 On Aug 26, 1999 this sequence version replaced gi:578561.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
 on the WORMPEP database can be found at

FEATURES

http://www.sanger.ac.uk/projects/C.elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1
 RP5-836J3 is from the library RPCL-5 constructed at the Roswell
 Park Cancer Institute by the group of Pletier de Jong. For further
 details see http://bacpac.med.buffalo.edu/
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP5-836J3.

source Location/Qualifiers
 1..143590
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="p22.1-31.1"
 /clone_id="RP5-836J3"
 /clone_1lb="RPCL-5"
 1160..1786
 /note="match: GSS: Em:B16380"
 complement(2068..2500)
 /note="match: GSS: Em:AQ089628"
 2554..3072
 /note="match: GSS: Em:AQ118130"
 complement(3824..4133)
 /note="match: GSS: Em:AQ089896"
 4017..4402
 /note="match: GSS: Em:AQ063212"
 4017..4312
 /note="match: GSS: Em:AQ059070"
 4023..4278
 /note="match: GSS: Em:AQ063209"
 6000..6045
 /note="2 copies 23 mer 100% conserved"
 6637..6906
 /note="ORSL repeat: matches 190..467 of consensus"
 9859..10375
 /note="MLTIF repeat: matches 3..536 of consensus"
 13112..13891
 /note="LTR17 repeat: matches 1..780 of consensus"
 15130..15173
 /note="22 copies 2 mer tg 93% conserved"
 15133..15172
 /note="4 copies 10 mer gtgtgtgtgt 95% conserved"
 23083..23118
 /note="18 copies 2 mer tg 80% conserved"
 23328..23713
 /note="WSTD repeat: matches 1..394 of consensus"
 26442..26509
 /note="134 copies 2 mer gt 66% conserved"
 complement(26847..27324)
 /note="match: GSS: Em:AQ835456"
 27344..27699
 /note="match: GSS: Em:AQ479571"
 31575..32054
 /note="match: STS: Em:AL009625"
 32094..32491
 /note="LTR2D repeat: matches 1..490 of consensus"
 32492..32857
 /note="match: GSS: Em:AQ140889"
 32521..32580
 /note="6 copies 10 mer gatagataga 80% conserved"
 32528..32587
 /note="15 copies 4 mer atag 96% conserved"
 32591..32635
 /note="LTR2D repeat: matches 505..553 of consensus"
 38429..38876
 /note="match: GSS: Em:AQ146760"
 39727..39756
 /note="6 copies 5 mer tgttt 96% conserved"
 40414..40453
 /note="20 copies 2 mer ca 97% conserved"
 42939..42989
 repeat_region

repeat_region	/note="MER91A repeat: matches95.152 of consensus43515.43610
misc_feature	/note="2 copies 48 mer 85% conserved"
misc_feature	/note="match(43654..44158)
misc_feature	/note="match: GSS: Em:AQ437799"
repeat_region	44168..44654
repeat_region	/note="match: GSS: Em:AQ64589"
repeat_region	45309..45340
misc_feature	/note="16 copies 2 mer tt 93% conserved"
repeat_region	45341..45660
repeat_region	/note="match: STS: Em:G24052"
repeat_region	45473..45610
repeat_region	/note="6 copies 23 mer 67% conserved"
repeat_region	46017..46066
repeat_region	/note="25 copies 2 mer ac 78% conserved"
repeat_region	46019..46066
repeat_region	/note="12 copies 4 mer aacac 79% conserved"
repeat_region	49244..49451
repeat_region	/note="MER5A repeat: matches 1..223 of consensus"
repeat_region	51847..51898
repeat_region	/note="13 copies 4 mer ctct 73% conserved"
repeat_region	51849..51898
misc_feature	/note="25 copies 2 mer ct 74% conserved"
repeat_region	complement(52909..53347)
misc_feature	/note="match: GSS: Em:AQ440464"
repeat_region	56014..56544
repeat_region	/note="MER6B repeat: matches 1..568 of consensus"
repeat_region	57696..57814
repeat_region	/note="MER6 repeat: matches 2..129 of consensus"
repeat_region	58171..58704
repeat_region	/note="MIT1 repeat: matches 3..548 of consensus"
misc_feature	66786..67150
repeat_region	/note="match: GSS: Em:AQ053163"
repeat_region	67192..67597
misc_feature	/note="MER7A repeat: matches 1..346 of consensus"
repeat_region	complement(68816..69390)
misc_feature	/note="match: GSS: Em:AQ549150"
repeat_region	69397..69432
repeat_region	/note="9 copies 4 mer taag 86% conserved"
misc_feature	69519..70015
repeat_region	/note="match: GSS: Em:AQ045852"
repeat_region	69911..69986
repeat_region	/note="19 copies 4 mer taat 65% conserved"
repeat_region	70512..71004
repeat_region	/note="MER4A repeat: matches 110..656 of consensus"
repeat_region	71303..71412
repeat_region	/note="MER4A repeat: matches 1..110 of consensus"
misc_feature	complement(73648..74145)
repeat_region	/note="match: GSS: Em:AQ458708"
repeat_region	75280..75333
repeat_region	/note="27 copies 2 mer tt 72% conserved"
repeat_region	76040..76115
repeat_region	/note="19 copies 4 mer aaga 97% conserved"
repeat_region	76046..76115
repeat_region	/note="35 copies 2 mer ga 75% conserved"
repeat_region	76047..76126
repeat_region	/note="8 copies 10 mer aagaagaa 76% conserved"
repeat_region	78923..79419
repeat_region	/note="MIT1 repeat: matches 67..541 of consensus"
repeat_region	80005..80032
repeat_region	/note="7 copies 4 mer atac 96% conserved"
repeat_region	82920..83118
repeat_region	/note="MER91A repeat: matches 2..195 of consensus"
repeat_region	83123..83171
repeat_region	/note="MER91A repeat: matches 138..186 of consensus"
repeat_region	83659..84155
repeat_region	/note="11P82 repeat: matches 5634..6155 of consensus"
repeat_region	87050..87201
repeat_region	/note="MER4A repeat: matches 1..148 of consensus"
repeat_region	87197..87556
repeat_region	/note="MER4A repeat: matches 394..728 of consensus"
misc_feature	91670..92052
misc_feature	/note="match: GSS: Em:AQ25888"

misc_feature	91670..92091	/note="match: GSS: Em:AQ279665"	
misc_feature	91670..92003	/note="match: GSS: Em:AQ278706"	
repeat_region	93141..94747	/note="WSTA-internal repeat: matches 1..1637 of consensus"	
repeat_region	94761..95183	/note="WSTB repeat: matches 1..426 of consensus"	
repeat_region	95891..96457	/note="LIMC1 repeat: matches 5219..5796 of consensus"	
repeat_region	96486..96839	/note="MER74A repeat: matches 165..524 of consensus"	
repeat_region	96940..97505	/note="LIMC1 repeat: matches 5791..6325 of consensus"	
repeat_region	100936..100937	/note="Alus repeat: matches 230..291 of consensus"	
repeat_region	105831..106139	/note="LOR1b repeat: matches 1..461 of consensus"	
repeat_region	106582..107055	/note="LIMB5 repeat: matches 5688..6174 of consensus"	
repeat_region	107066..107248		
Query Match	72.0%	Score 359.8; DB 9;	Length 143590;
Best Local Similarity	86.3%	Pred. No. 4.8e-108;	
Matches 414;	Conservative	0; Mismatches 52;	Indels 3; Gaps 2;

Query Match	72.0%	Score 359.8	DB 9	Length 143590
Best Local Similarity	88.3%	Prod. No. 4.8e-108		
Matches 414	Conservative 0	Mismatches 52	Indels 3	Gaps 2
QY 34	AAAAGCCTGAGATATACAGCAATTATCTTGGCACTGAGAGACAGGACTAGCTGATTTCC	93		
Db 13925	AAAATGATAGAGCTACCAAGTATTGAGACACTACTGAGAGATAGGACTAGCTGATTTCC	13866		
QY 94	TAGGCGCACTAAGATATCCCTTAAGCCTAGCTGGGAAGGTGACCAAGTCCACCTTTAAACAC	153		
Db 13865	TAGGCGCACTAAATATCCCTTAAGCCTAGTTGGGAAGGTGACCGCATCCACCTTTAAACAT	13806		
QY 154	GGGGCTTGCACTTAGCTCAACCTGACCAATCAAGAGGCTCACTAAATGCTAATTAGG	213		
Db 13805	GGGGCTTGCACTTAGCTCAACCCCAACCAATCAAGAGGCTCACTAAATGCTAATTAGG	13746		
QY 214	CAAGAAGCAGAGGTAAAGAAATAGCCATCATATTCTCTGAGAGCAGCAGAGAGGGA	273		
Db 13745	CAAAAACAGAGGTAGAGAAATAGCCATCATATTCTCTGAGAGCAGCAGTGTAGGGA	13686		
QY 274	CAACAATCGGGATATAAACCAGGCAATTGAGCTGGCAACAGACAGCCCCCTTTGGGTCC	333		
Db 13685	CAAGAGATCGGGATATAAACCCAGGCAATTGAGCGGGCAACGGCAG-CCCCCTTTGGGTCC	13627		
QY 334	CTTCCCTTTGATAGGAG--CTGTTTTCATCTTATTTCATCTCTATTAATCTTGCACCTG	391		
Db 13626	CTTCCCTTTGATAGGAGCTCTGTTTTCACCTCTATTTCATCTCTATTAATCTTGCACCTG	13567		
QY 392	CACCTCTTGCTGCATGTTTCTTAAGGCTGAGCGTCTTGGCTCAACGCGCACCACT	451		
Db 13566	CACCTCTTGCTGCATGTTTCTTAAGGCTGAGCGTCTTGGCTCAACGCGCACCACT	13507		
QY 452	GCTGTTTGGCCACACCGCAGACTGCGCTGACTCCCATCCCTCTGAT	500		
Db 13506	GCTGTTTGGCCACCGCTCCACGACCGGTACTGACTTCATCCCTCTGAT	13458		
RESULT 10	AC103595	119406 bp	DNA	linear HTG 29-NOV-2001
AC103595/c				
LOCUS	Homo sapiens chromosome 1 clone RP11-415P10, WORKING DRAFT			
DEFINITION	SEQUENCE, 10 unordered pieces.			
ACCESSION	AC103595			
VERSION	AC103595.1			
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 119406)			
AUTHORS	Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.			

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Unpublished
2 (bases 1 to 119406)
Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
Direct Submission
Submitted (29-NOV-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, US.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
----- Project Information
Center project name: chr-1
Center clone name: RP11-415P10 (sc0362)

----- Summary Statistics
Sequencing vector: plasmid; L03752; 100% of reads
Chemistry: Dye-terminator ET; 32% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 109646 bases; at least Q40
Consensus quality: 115473 bases; at least Q30
Consensus quality: 117774 bases; at least Q20
Insert size: 194800; 40.1% error; agarose-fp
Insert size: 118506; sum-of-contigs
Quality coverage: 5.7x in Q20 bases; agarose-fp
Quality coverage: 9.3x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 4564: contig of 4564 bp in length
* 4565 4664: gap of unknown length
* 4665 12930: contig of 8266 bp in length
* 12931 13030: gap of unknown length
* 13031 20784: contig of 7754 bp in length
* 20785 20884: gap of unknown length
* 20885 29093: contig of 8209 bp in length
* 29094 29193: gap of unknown length
* 29194 40511: contig of 11316 bp in length
* 40512 40611: gap of unknown length
* 40612 52409: contig of 11796 bp in length
* 52410 52509: gap of unknown length
* 52510 64775: contig of 12266 bp in length
* 64776 64875: gap of unknown length
* 64876 83398: contig of 18523 bp in length
* 83399 83498: gap of unknown length
* 83499 100056: contig of 16558 bp in length
* 100057 100156: gap of unknown length
* 100157 119406: contig of 19250 bp in length.

FEATURES
source
1. 119406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-415P10"
/clone_1b="RP11 human BAC library 11"
1. 4564
/note="assembly_name:Contig39"
misc_feature
4665..12930
/note="assembly_name:Contig30"
misc_feature
13031..20784
/note="assembly_name:Contig31"
misc_feature
20885..29093
/note="assembly_name:Contig32"
misc_feature
29194..40511
/note="assembly_name:Contig33"
misc_feature
40612..52409

misc_feature
52510..64775
/note="assembly_name:Contig35"
misc_feature
64876..83398
/note="assembly_name:Contig36"
misc_feature
83499..100056
/note="assembly_name:Contig37"
misc_feature
100157..119406
/note="assembly_name:Contig38"

BASE COUNT 35588 a 23913 c 23787 g 35137 t 981 others
ORIGIN

Query Match 71.6%; Score 358.2; DB 2; Length 119406;
Best Local Similarity 88.1%; Pred. No. 1.6e-107;
Matches 413; Conservative 0; Mismatches 53; Indels 3; Gaps 2;

QY 34 AAAAGCGCTGAGATTCAGACATTTATCTTGCACTGAGAGAGAGACTAGCTGATTTCC 93
DB 37607 AAAATGTACGACGCTACGACTATTGAGACATCTGAGAGATGAGACTGATTTCC 37548

QY 94 TAGGCCGACTAAGATTCCTTAAGCTTAGCTGGAGGTGACACGCTTCACTTTAAAC 153
DB 37547 TAGCCGACCTAAATTCCTTAAGCTTAGCTGGAGGTGACCGCATCCTTTAAACAT 37488

QY 154 GGGGCTTGCACTTAGCTTCACCTGACCAATGACAGAGCTCACTAAATGCTTAATAG 213
DB 37487 GGGGCTTGCACTTAGCTTCACCTGACCAATGACAGAGCTCACTAAATGCTTAATAG 37428

QY 214 CAAGAAGAGAGGTAAGAATAGCCCAATCATCTATTCCTGAGAGAGAGAGAGGA 273
DB 37427 CAAGAAGAGAGGTAAGAATAGCCCAATCATCTATTCCTGAGAGAGAGAGAGGA 37368

QY 274 CAACATCGGATTAATTAACCGAGCATTCGAGCTGGCAAGACGCCCTTTGGGCTC 333
DB 37367 CAAGATCGGATTAATTAACCGAGCATTCGAGCTGGCAAGACGCCCTTTGGGCTC 37309

QY 334 CTTCCTTTGATGAGG--CTGTTTACATGCTATTTCATCTATTAACTTGCAACTG 391
DB 37308 CTTCCTTTGATGAGGCTCTGTTTACATGCTATTTCATCTATTAACTTGCAACTG 37249

QY 392 CACCTTCTGGTCCAGTTTCTTACGCTGAGCTGAGCTTTTCTGACCGTCACCACT 451
DB 37248 CACCTTCTGGTCCGTTTCTTACGCTGAGCTGAGCTTTTCTGACCGTCACCACT 37189

QY 452 GCTGTTTCCACGACCGCAGACTGCGCTGACTCCCATCTCTGTGAT 500
DB 37188 GCTGTTTCCACGCTCCGACGACCGCTGACTGCTTCATCTCTGTGAT 37140

RESULT 11
AL353584 0/c
WPCOMMENT
Sequence split into 6 fragments LOCUS AL353584 Accession AL353584
Fragment Name Begin End
AL353584_0 1 11000
AL353584_1 10001 21000
AL353584_2 20001 31000
AL353584_3 30001 41000
AL353584_4 40001 51000
AL353584_5 50001 536214
LOCUS AL353584 536214 bp DNA linear HTG 19-JUL-2001
DEFINITION Homo sapiens chromosome X clone RP13-185F3, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL353584
VERSION AL353584.11 GI:15020381
KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 536214)
AUTHORS Wilson,S.
TITLE Direct Submission

Query Match 71.6%; Score 357.8; DB 9; Length 169029;
Best Local Similarity 92.0%; Pred. No. 2 3e-107;
Matches 400; Conservative 0; Mismatches 32; Indels 3; Gaps 2;

QY 68 TGAGAGACAGACTGATGATTTCTTAGCCGACCTAAGAAATCCCTTAAGCTTACTGAGGA 127
Db 22071 TGAGAGACAGACTGATGATTTCTTAGCCGACCTAAGAAATCCCTTAAGCTTACTGAGGA 22012
QY 128 AGGAGACAGCTGACCTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAATCA 187
Db 22011 AGGAGACAGCTGACCTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAATCA 21952
QY 188 GAGAGCTCACTAAATGCTAATTAAGACAGAGAGTAAGAAATAGCCATCATCT 247
Db 21951 GAGAGCTCACTAAATGCTAATTAAGACAGAGAGTAAGAAATAGCCATCATCT 21892
QY 248 ATTGCTGAGAGCAGACAGAGAGGACACAAATGGGATATTAACCCAGGACTTGAGCT 307
Db 21891 ATTGCTGAGAGCAGACAGAGAGGAGACAAATGGGATATTAACCCAGGACTTGAGCT 21832
QY 308 GGGACAGACAGACCCCTTTGGGTCCTTCCCTTGGATGGAG--CTGTTTCATGCTA 365
Db 21831 GGGACAGAGCA-ACCCCTTTGGGTCCTTCCCTTGGATGGAGAGCTGTTTCATGCTA 21773
QY 366 TTGACTCTAATTAATCTTGCACTGCTCTGCTGCTCAATGTTTCTTAAGGCTGAGC 425
Db 21772 TTGACTCTAATTAATCTTGCACTGCTCTGCTGCTCAATGTTTCTTAAGGCTGAGC 21713
QY 426 TGAAGTTTGTCTACCGTCCACCACTGCTGTTTGCCACACCGGAGACTGCGCTGACT 485
Db 21712 TGAAGTTTGTCTACCGTCCACCACTGCTGTTTGCCACACCGGAGACTGCGCTGACT 21653
QY 486 CCCATCCCTCTGGAT 500
Db 21652 CCCATCCCTCTGGAT 21638

RESULT 13
AL392173/c 170746 bp JMA linear PRI 05-APR-2001
LOCUS Human DNA sequence from clone RP11-158M9 on chromosome Xq26.1-27.1
DEFINITION Contains STS and GSSs, complete sequence.
ACCESSION AL392173
VERSION AL392173.8 GI:11064224
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 170746)
Bray-Allen S.
Direct Submission
Submitted (07-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humgery@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
On Oct 31, 2000 this sequence version replaced gi:10800700.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
This sequence is the entire insert of clone RP11-158M9 This
sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP11-158M9 is from
the library RPI1-11.1 constructed by the group of Pieter de Jong.
For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.

FEATURES

source
Location/Qualifiers

1..170746
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q26.1-27.1"
/clone="RP11-158M9"
/clone_1fb="RPCT-11.1"
688..1038
/note="match: GSS: Em:AQ314307"
2318..2343
/note="13 copies 2 mer ca 100% conserved"
7251..7937
/note="L1PA6 repeat: matches 4764..5447 of consensus"
7938..8568
/note="L1PA6 repeat: matches 5507..6139 of consensus"
9828..10340
/note="match: GSS: Em:A2368366"
14268..14619
/note="MER34 repeat: matches 4..301 of consensus"
20566..21107
/note="MLR2B repeat: matches 1..448 of consensus"
27734..27963
/note="match: STS: Em:HS210VD10"
27963..27998
/note="18 copies 2 mer ac 100% conserved"
27967..28062
/note="match: STS: Em:HS210VD10"
29226..29569
/note="TRH1C repeat: matches 1..371 of consensus"
29570..30022
/note="TRH1C internal repeat: matches 1..445 of consensus"
30335..31434
/note="TRH1C-internal repeat: matches 445..1580 of consensus"
31435..31807
/note="TRH1C repeat: matches 1..371 of consensus"
35205..35669
/note="match: GSS: Em:AQ493428"
37464..38246
/note="L1R17 repeat: matches 1..780 of consensus"
39054..39117
/note="13 copies 2 mer gt 92% conserved"
41432..41546
/note="L1R29 repeat: matches 516..619 of consensus"
41877..42007
/note="L1R29 repeat: matches 3..149 of consensus"
51113..51140
/note="14 copies 2 mer tc 89% conserved"
51268..51318
/note="H15 repeat: matches 30..76 of consensus"
54978..55013
/note="18 copies 2 mer ca 91% conserved"
58352..59188
/note="TRH1B-INTERNAL repeat: matches 1..842 of consensus"
61106..61272
/note="W5TB-internal repeat: matches 15..186 of consensus"
61578..61999
/note="W5TB-internal repeat: matches 186..585 of consensus"
62003..62098
/note="W5T-INTERNAL repeat: matches 594..689 of consensus"
63045..63698


```

repeat_region /note="HERV16 repeat: matches 4. .664 of consensus"
63927. .64774 /note="HERV16 repeat: matches 716. .1577 of consensus"
repeat_region 64789. .64832 /note="22 copies 2 mer aa 81% conserved"
repeat_region 64916. .65612 /note="HERV16 repeat: matches 1817. .2518 of consensus"
repeat_region 65613. .66886 /note="THER1-INTERNAL repeat: matches 124. .1401 of consensus"
misc_feature 66475. .66484 /note="1372 bp of IS186 transposon (X03123) removed here. This sequence represents the duplicated flanking sequence of the IS186."
70601. .70772 /note="match: GSS: Em:AQ077330"
70734. .71229 /note="match: GSS: Em:AQ628837"
misc_feature 80000. .80512 /note="match: GSS: Em:AQ284023"
misc_feature match: STS: Em:G61704"
complement(81135. .81789)
/note="match: GSS: Em:AQ240981"
81193. .81655 /note="match: GSS: Em:AQ704174"
83625. .89762 /note="LIPAS repeat: matches 11. .6145 of consensus"
91802. .91895 /note="LIPAS repeat: matches 6053. .6146 of consensus"
94922. .101052 /note="LIPAS repeat: matches 11. .6145 of consensus"
102779. .102828 /note="LIPAS repeat: matches 11. .6145 of consensus"
complement(103577. .104157)
/note="25 copies 2 mer ac 82% conserved"
105177. .105697 /note="match: GSS: Em:AQ775154"
105842. .106341 /note="match: GSS: Em:B63384"
10797. .108270 /note="match: GSS: Em:AQ522879"
116826. .116918 /note="MER39 repeat: matches 13. .521 of consensus"
complement(122839. .123304)
/note="MER31A repeat: matches 1. .94 of consensus"
126653. .126696 /note="match: GSS: Em:AQ002183"
127908. .127950 /note="22 copies 2 mer ac 93% conserved"
/note="LIPAS repeat: matches 6103. .6145 of consensus"
complement(129916. .130692)
/note="match: GSS: Em:AQ288149"
complement(130168. .130745)
/note="match: GSS: Em:AQ293756"
complement(131672. .132168)
/note="match: GSS: Em:AQ466443"
131821. .131846 /note="13 copies 2 mer ta 96% conserved"
132163. .132600 /note="match: GSS: Em:AQ829322"
133493. .133664 /note="LIPAS repeat: matches 30. .202 of consensus"
136095. .136245 /note="LIPAS repeat: matches 6142. .6289 of consensus"
complement(140979. .141356)
/note="match: STS: Em:HSC28H08"
146003. .146304 /note="LIPAS repeat: matches 1. .310 of consensus"
146351. .146421 /note="LIPAS repeat: matches 448. .516 of consensus"
157021. .157123 /note="MER34 repeat: matches 76. .173 of consensus"
159023. .159689 /note="HERV16 repeat: matches 562. .1217 of consensus"

```

```

repeat_region 15978. .160033 /note="HERV16 repeat: matches 1217. .1272 of consensus"
repeat_region 164301. .165134 /note="LIPAS repeat: matches 1841. .3688 of consensus"
repeat_region 166458. .166509 /note="26 copies 2 mer ga 73% conserved"
repeat_region 166568. .169037 /note="LIPAS repeat: matches 3681. .6148 of consensus"
misc_feature 169389. .169794 /note="match: GSS: Em:AQ092454"
misc_feature complement(170139. .170746)
misc_feature /note="match: GSS: Em:AQ389950"
misc_feature complement(170286. .170416)
/note="match: GSS: Em:AQ835692"
BASE COUNT 53940 a 32801 c 34298 g 49707 t
ORIGIN

```

```

Query Match 71.6%; Score 357.8; DB 9; Length 170746;
Best Local Similarity 92.0%; Pred.No.2.3e-107;
Matches 400; Conservative 0; Mismatches 32; Indels 3; Gaps 2;

QY 68 TGAGAGACAGACTAGCTGATTTCTAGGCGGACGACTAAGATCCCTAAGCTAGCTGGGA 127
Db 38246 TGAGAGACAGACTAGCTGATTTCTAGGCGGACGACTAAGATCCGAGCTAGCTGGGA 38187

QY 128 AGGTGACACAGCTCACCCTTTAAACAGGGGCTTGCACTTAGCTCACCTGACCAATCA 187
Db 38186 AGGTGACAGCTCACCCTTTAAACAGGGGCTTGCACTTAGCTCACCTGACCAATCA 38127

QY 188 GAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATGACCAATCATCT 247
Db 38126 GAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATGACCAATCATCT 38067

QY 248 ATTGCGTAGAGAGACAGACGAGGAGCAACATCGGATTTAAACCAAGCATTCGAGCT 307
Db 38066 ATTGCGTAGAGAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 38007

QY 308 GGGACAGACAGCCCCCTTTGGGCTCCCTTCTTGTATGGAG--CTGTTTCATGCTA 365
Db 38006 GGGACAGGCA-ACCCCTTTGGGCTCCCTTCTTGTATGGAGCTGTTTCATGCTG 37948

QY 366 TTTCATCTAATTAATCTTGCAACTGCACTCTTGTGCTCATGTTTCTTAAGCTGAGC 425
Db 37947 TTTCATCTAATTAATCTTGCAACTGCACTCTTGTGCTCATGTTTCTTAAGCTGAGC 37888

QY 426 TGAGCTTTTGTCAACCGTCCACAGCTGCTTTGGCAACACCGGACACCTGCGTGAAT 485
Db 37887 TGAGCTTTTGTCAACCGTCCACAGCTGCTTTGGCCGCGTCCGACACCGGCTGACT 37828

QY 486 CCCATCCCTCTGGAT 500
Db 37827 TCCATCCCTCCAGAT 37813

```

```

RESULT 14
AC093531/c 163803 bp DNA linear PRI 16-NOV-2001
LOCUS Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
DEFINITION AC093531
VERSION AC093531.2 GI:16945981
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
AUTHORS Unpublished
RECORD 2 (bases 1 to 163803)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission.
JOURNAL Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint

```


Db 171735 TCGAGCTGAGCTTTGCTCGCCATCCACACAGCTGTTTGCCGCTTGACAGACCCGCTG 171794
Qy 480 CTGACTCCCATCCCTCTGGAT 500
Db 171795 CTGACTCCCATCCCTCTCGAT 171815

Search completed: April 19, 2003, 13:48:52
Job time : 2068.07 secs

XX Claim 3; Fig 1; 225bp; French.
XX
XX The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neopathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;
Query Match 100.0%; Score 500; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 5,1e-164;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAACGCTGAGATACAGCAATTATC 60
DB 1 CCTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAACGCTGAGATACAGCAATTATC 60
QY 61 TTGCACTGAGAGACAGAGCTAGCTGATTTCTCTAGG TCGACTAAGAACTCTAAGCCTA 120
DB 61 TTGCACTGAGAGACAGAGCTAGCTGATTTCTCTAGG TCGACTAAGAACTCTAAGCCTA 120
QY 121 GCTGGGAAGGTACACGCTCCACTTTAAACAGGGGGTTTGCACTTAGCTACACCTGA 180
DB 121 GCTGGGAAGGTACACGCTCCACTTTAAACAGGGGGTTTGCACTTAGCTACACCTGA 180
QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGGTAAAGAAATAGCCA 240
DB 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGGTAAAGAAATAGCCA 240
QY 241 ATCATCTATTGGCTGAGACAGACAGAGGAGCAACATCGGAGATATAACCCGAGCAT 300
DB 241 ATCATCTATTGGCTGAGACAGACAGAGGAGCAACATCGGAGATATAACCCGAGCAT 300
QY 301 TCGAGCTGGCAACAGAGAGCCCTTTGGGCTCCCTTTGTATGGAGCTGTTTTCA 360
DB 301 TCGAGCTGGCAACAGAGAGCCCTTTGGGCTCCCTTTGTATGGAGCTGTTTTCA 360
QY 361 TGGTATTGCTCTATTAAATCTTGGCACTGCACTGTTTGGTCCATGTTTCTTAAGGCT 420
DB 361 TGGTATTGCTCTATTAAATCTTGGCACTGCACTGTTTGGTCCATGTTTCTTAAGGCT 420
QY 421 CGAGCTGAGCTTTTGGTCCACGTCACCACTGCTGTTTCCACACCGGAGACCTGGCCG 480
DB 421 CGAGCTGAGCTTTTGGTCCACGTCACCACTGCTGTTTCCACACCGGAGACCTGGCCG 480
QY 481 TGAATCCCATCCCTCTGGAT 500
DB 481 TGAATCCCATCCCTCTGGAT 500
RESULT 2
ABL61744
ID ABL61744 standard; DNA; 56093 BP.
XX
XX ABL61744;
XX
XX 15-MAY-2002 (first entry)
XX
XX Colon adenocarcinoma related gene sequence S3Q ID NO:81.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytostatic; gene therapy; antineoplastic; Wilms' tumor; adenocarcinoma;
KW

XX gene; ds.
XX
XX Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US10838.
XX
XX 05-JUN-2000; 2000US-209472P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-231133P.
XX 18-SEP-2000; 2000US-231617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 20-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 25-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235863P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.
XX 28-SEP-2000; 2000US-236109P.
XX 28-SEP-2000; 2000US-236111P.
XX 29-SEP-2000; 2000US-236842P.
XX 29-SEP-2000; 2000US-236891P.
XX 02-OCT-2000; 2000US-237172P.
XX 02-OCT-2000; 2000US-237173P.
XX 02-OCT-2000; 2000US-237278P.
XX 02-OCT-2000; 2000US-237294P.
XX 02-OCT-2000; 2000US-237295P.
XX 02-OCT-2000; 2000US-237315P.
XX 03-OCT-2000; 2000US-237425P.
XX 03-OCT-2000; 2000US-237598P.
XX 03-OCT-2000; 2000US-237604P.
XX 03-OCT-2000; 2000US-237606P.
XX 03-OCT-2000; 2000US-237608P.
XX 01-NOV-2000; 2000US-244867P.
XX 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Sopet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 81; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (i)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL6164
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (i) has cytostatic

CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.

SQ Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 other;

Query Match 100.0%; Score 500; DB 24; Length 56093;

Best Local Similarity 100.0%; Pred. No. 1.3e-163; Mismatches 0; Gaps 0;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGTGGGGCGGGCTTCCTTCTGGGATGAGGGGCAAAAGCGCTGAGATACAGCAATTATC 60
DB 28001 CCTGTGGGGCGGGCTTCCTTCTGGGATGAGGGGCAAAAGCGCTGAGATACAGCAATTATC 28060
QY 61 TTGCAACTGAGACAG 120
DB 28061 TTGCAACTGAGACAG 28120
QY 121 GCTGGGAGAGTGAACAAGTCCACTTTAAACAGGGGCTTGAACCTAGCTCAACCTGA 180
DB 28121 GCTGGGAGAGTGAACAAGTCCACTTTAAACAGGGGCTTGAACCTAGCTCAACCTGA 28180
QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 28181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 28240
QY 241 ATATCTATTGCTCTGAG 300
DB 28241 ATATCTATTGCTCTGAG 28300
QY 301 TCAGAGTGGCAAG 360
DB 28301 TCAGAGTGGCAAG 28360
QY 361 TGCTATTTCATCTATTAAATCTTGAAGTGAAGTCTTGTGTCATGTTTTCAGGGCT 420
DB 28361 TGCTATTTCATCTATTAAATCTTGAAGTGAAGTCTTGTGTCATGTTTTCAGGGCT 28420
QY 421 CGAGCTGAGCTTTTGTCTCAACCTGTCACACTGCTGTTGGCAACGAGAGAGAGAGAGAGAG 480
DB 28421 CGAGCTGAGCTTTTGTCTCAACCTGTCACACTGCTGTTGGCAACGAGAGAGAGAGAGAG 28480
QY 481 TGACTCCCATCCCTCTGGAT 500
DB 28481 TGACTCCCATCCCTCTGGAT 28500

RESULT 3
ABN97946
ID ABN97946 standard; DNA; 711 BP.

XX ABN97946;

XX 01-AUG-2002 (first entry)

XX Human retroviral sequence R1.

XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

XX Multiple sclerosis; ds.

XX Human retrovirus.

XX MO967395-A1.
XX 29-DEC-1999.

XX

PF 23-JUN-1999; 99WO-FR01513.

XX 23-JUN-1998; 98FR-0007920.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Alliel PM, Perin J, Rieger F;

XX WPI; 2000-160587/14.

PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,

PT used for diagnosis, treatment and prevention of autoimmune and

PS Claim 3; Fig 3; 225pp; French.

CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.

SQ Sequence 711 BP; 170 A; 204 C; 162 G; 175 T; 0 other;

Query Match 87.0%; Score 435; DB 21; Length 711;

Best Local Similarity 100.0%; Pred. No. 7.1e-142; Mismatches 0; Gaps 0;

Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ACTGAGAGACAGAGACTAGTGAATTCCTAGGCGCACTAAGATCCCTAGCCTAGCTGG 125
DB 1 ACTGAGAGACAGAGACTAGTGAATTCCTAGGCGCACTAAGATCCCTAGCCTAGCTGG 60
QY 126 GAGGTGACCAAGTCCACTTTAAACAGGGGCTTGAACCTAGCTCAACCTGACCAAT 185
DB 61 GAGGTGACCAAGTCCACTTTAAACAGGGGCTTGAACCTAGCTCAACCTGACCAAT 120
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245
DB 121 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 246 CTATTGCTGAGAGACAG 305
DB 181 CTATTGCTGAGAGACAG 240
QY 306 CTGGCAACAG 365
DB 241 CTGGCAACAG 300
QY 366 TTTCACCTATTAAATCTTGAAGTGAAGTCTTGTGTCATGTTTTCAGGGCTGAGC 425
DB 301 TTTCACCTATTAAATCTTGAAGTGAAGTCTTGTGTCATGTTTTCAGGGCTGAGC 360
QY 426 TGAGCTTTTGTCTCAACCTGTCACACTGCTGTTGGCAACGAGAGAGAGAGAGAGAGAG 485
DB 361 TGAGCTTTTGTCTCAACCTGTCACACTGCTGTTGGCAACGAGAGAGAGAGAGAGAGAG 420
QY 486 CCATCCCTCTGGAT 500
DB 421 CCATCCCTCTGGAT 435

RESULT 4
AAS68626
ID AAS68626 standard; cDNA; 7466 BP.

XX

AC AAS68626;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4430.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
FI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG04439.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 4430; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1154 T; 1 other;

Query Match 82.7%; Score 413.4; LB 23; Length 7466;
Best Local Similarity 98.2%; Pred. No. 1e-133;
Matches 429; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 66 ACTGAGACGACGACTGATTTCTTCCAGGCGCAGCTAGAGTCCCTAGGCTGAGG 125
DB 3716 ACTGAGACGACGACTGATTTCTTCCAGGCTGATTTAGAGTCCCTAGGCTGAGG 3775
QY 126 GAAGGTGACGACGCTCACTTTAAACACGCGGCTTGCAATTAGCTCACACCTGACCAAT 185
DB 3776 GAAGGTGACGACGCTCACTTTAAACACGCGGCTTGCAATTAGCTCACACCTGACCAAT 3835
QY 186 CAGAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGGTGTAAGAAATAGCCATCAT 245
DB 3836 CAGAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGGTGTAAGAAATAGCCATCAT 3895

QY 246 CTATTGCTGAGACAGCAGAGGACCAATCGGATATTAACCCAGGCAATTCGAG 305
DB 3896 TTATTGCTGAGACAGCAGAGGACCAATGATCGGATATTAACCCAGGCAATTCGAG 3955
QY 306 CTGCAACAGCAGGCCCCCTTTGGGTCCTTCCCTTTGATGGAG--CTGTTTCATGC 363
DB 3956 CTGCAACAGCAGGCCCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTCATGC 4015
QY 364 TATTCACTATTAATCTTGGCACTGCTTGGTCATGTTCTTAACGGCTCGA 423
DB 4016 TATTCACTATTAATCTTGGCACTGCTTGGTCATGTTCTTAACGGCTCGA 4075
QY 424 GCTGAGCTTTTGTACCGTCACACCACTGCTGTTGGCACCAGCAGACCTGCGCTGA 483
DB 4076 GCTGAGCTTTTGTACCGTCACACCACTGCTGTTGGCACCAGCAGACCTGCGCTGA 4135
QY 484 CTCCCATCCTCTTGAGAT 500
DB 4136 CTCCCATCCTCTTGAGAT 4152

RESULT 5
AAX25669
ID AAX25669 standard; cDNA to mRNA; 783 BP.

AAX25669;

XX 21-MAY-1999 (first entry)

DE Human endogenous retrovirus W long terminal repeat region.

KM Clone; human endogenous retrovirus; genome; autoimmune disease;
KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

OS Human endogenous retrovirus.

PN WO9902696-A1.

PD 21-JAN-1999.

PP 06-JUL-1998; 98WO-FR01442.

PR 07-JUL-1997; 97FR-0008815.

PA (IMMR) BIO MERIEUX.

PI Beeseme F, Blond JL, Boulton O, Mallet F, Mandrand B;

DR WPI; 1999-120897/10.

PT New nucleic acid sequences from human endogenous retrovirus-W -
PT expressed exclusively in placenta and useful in diagnosis and
PT therapy of autoimmune disease, and abnormal or failed pregnancy

PS Claim 1; Page 83; 106pp; French.

XX This sequence represents the long terminal repeat (LTR) region of the
XX human endogenous retrovirus (HERV) W genome. The nucleic acids, their
XX fragments or peptides encoded by them are markers of autoimmune disease
XX (e.g. multiple sclerosis, rheumatoid polyarthritis, disseminated lupus
XX erythematosus, insulin-dependent diabetes and related pathologies) and
XX of abnormal or unsuccessful pregnancy and can be used as chromosomal
XX markers for susceptibility to these conditions, or proximity markers
XX of genes associated with this susceptibility.

SQ Sequence 783 BP; 173 A; 213 C; 166 G; 180 T; 51 other;

Query Match 82.0%; Score 410; DB 20; Length 783;
Best Local Similarity 93.8%; Pred. No. 4.5e-133;
Matches 408; Conservative 24; Mismatches 1; Indels 2; Gaps 1;

XX MO9823755-A1.
XX 04-JUN-1998.
XX 26-NOV-1997; 97WO-IB01482.
XX 26-NOV-1996; 96US-0756429.
XX (INMR) BIO MERIEUX.
XX Bedin F, Beeseme F, Jolivet-Reynaud C, Komurian-Pradel F,
XX Mandrand B, Paranhos-Baccala G, Perron J;
XX WPI; 1998-332732/28.
XX P-PSDB; AAW71069.
XX New nucleic acid from retroviruses - useful for diagnosis,
XX prevention and treatment of, e.g. multiple sclerosis
XX
XX Disclosure; Pages 187-188; 286pp; English.
XX
XX The present sequence represents a multiple sclerosis (MS) associated
XX retrovirus (MSRV) genomic fragment used in the method of the
XX invention. The invention provides complete or partial genomic
XX sequences of the MSRV-1 pol gene, gag gene and env gene, and
XX polypeptides encoded by these genes. The invention also provides
XX antibodies raised against the polypeptides. The genomic sequences,
XX polypeptides and antibodies are also claimed useful for diagnosing
XX infection by MS and rheumatoid arthritis-associated viruses, and also
XX for prevention and treatment of infection with these viruses.
XX
XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 69.6%; Score 348.2; DB 19; Length 1329;
Best Local Similarity 90.2%; Pred. No. 3e-111;
Matches 394; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

QY 66 ACTGAGACAGAGACTAGCTGATTTCTTAGGCGACCAAGAAATCCCTAAGCCTAGCTGG 125
DB 547 ACTGAGACAGAGACTAGCTGATTTCTTAGGCTAGCTAAGAAATCCCAAGCCTAAGCTGG 606
QY 126 GAAGGTGACACAGCTACCTTTAAACAGGGGGCTTGCACTTAGCTCACACCTGACCAAT 185
DB 607 GAAGGTGACCGCATCATCTTTAAACATGGGGCTTGCACTTAGCTCACACCGACCAAT 666
QY 186 CAGAGAGCTCACTAAATGCTAATTAAGGCAACAGAGAGTAAAGAAATAGCCATCAT 245
DB 667 CAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGAGGTAAAGCAATAGCCATCAT 726
QY 246 CTATTGCTGAGAGACAGACAGAGGAGCAACAATCGGATATTAACCCAGGCAATTGCG 305
DB 727 CTATTGCTGAGAGACAGACGGGAGAGACAAAGATTGGATTTAACTAGGCAATTCAG 786
QY 306 CTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTTGATGGAG--CTGTTTTCATGC 363
DB 787 CCAGCAACAGCA-ACCCCTTTGGGTCCCTCCCATTTGATGGAGCTGTTTTCATGC 845
QY 364 TATTTCACCTAATTAATTTGCACTGCTTTGATTCATGTTTCTTAAGGCTCGA 423
DB 846 TATTTCACCTAATTAATGATGCACTGCTTTGATTCGTTTCTTAAGGCTCGA 905
QY 424 GGTGAGCTTTTGTCTACCGTCCACCACTGCTGTTGGCCCAACCGGAGACTGCGGCTGA 483
DB 906 GGTGAGCTTTTGTGCGCATCCACCACTGCTGTTGGCCCGGTACAGACCGGCTGCTGA 965
QY 484 CTCCCATCCCTCTGGAT 500
DB 966 CTTCATCCCTTTGGAT 982

RESULT 8
AAK29704

ID AAK29704 standard; DNA; 1329 BP.
XX
XX AAK29704;
AC 08-JUN-1999 (first entry)
XX
XX
XX Clone 5M6 from MSRV-1.
DE
XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
XX rheumatoid polyarthritis; ss.
XX
XX Multiple sclerosis related virus type 1.
XX
XX FR2765588-A1.
XX
XX 08-JAN-1999.
XX
XX
XX 07-JUL-1997; 97FR-0008816.
XX
XX 07-JUL-1997; 97FR-0008816.
XX
XX 07-JUL-1997; 97FR-0008816.
XX
XX (INMR) BIO MERIEUX.
XX
XX WPI; 1999-098275/09.
XX P-PSDB; AAW9554.
XX
XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
XX multiple sclerosis or rheumatoid polyarthritis
XX
XX Claim 1; Page 39-40; 83pp; French.
XX
XX This sequence represents clone 5M6 from a novel multiple sclerosis
XX related virus type 1 (MSRV1). The sequence can be used in diagnostic,
XX prophylactic or therapeutic compositions to inhibit expression of a
XX multiple sclerosis related virus and/or virus associated with
XX rheumatoid polyarthritis.
XX
XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 69.6%; Score 348.2; DB 20; Length 1329;
Best Local Similarity 90.2%; Pred. No. 3e-111;
Matches 394; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

QY 66 ACTGAGACAGAGACTAGCTGATTTCTTAGGCGACCAAGAAATCCCTAAGCCTAGCTGG 125
DB 547 ACTGAGACAGAGACTAGCTGATTTCTTAGGCTAGCTAAGAAATCCCAAGCCTAAGCTGG 606
QY 126 GAAGGTGACACAGCTACCTTTAAACAGGGGGCTTGCACTTAGCTCACACCTGACCAAT 185
DB 607 GAAGGTGACCGCATCATCTTTAAACATGGGGCTTGCACTTAGCTCACACCGACCAAT 666
QY 186 CAGAGAGCTCACTAAATGCTAATTAAGGCAACAGAGAGTAAAGAAATAGCCATCAT 245
DB 667 CAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGAGGTAAAGCAATAGCCATCAT 726
QY 246 CTATTGCTGAGAGACAGACAGAGGAGCAACAATCGGATATTAACCCAGGCAATTGCG 305
DB 727 CTATTGCTGAGAGACAGACGGGAGAGCAAGATTGGATATTAACCTAGGCAATTCAG 786
QY 306 CTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTTGATGGAG--CTGTTTTCATGC 363
DB 787 CCAGCAACAGCA-ACCCCTTTGGGTCCCTCCCATTTGATGGAGCTGTTTTCATGC 845
QY 364 TATTTCACCTAATTAATTTGCACTGCTTTGATTCATGTTTCTTAAGGCTCGA 423
DB 846 TATTTCACCTAATTAATGATGCACTGCTTTGATTCGTTTCTTAAGGCTCGA 905
QY 424 GGTGAGCTTTTGTCTACCGTCCACCACTGCTGTTGGCCCAACCGGAGACTGCGGCTGA 483
DB 906 GGTGAGCTTTTGTGCGCATCCACCACTGCTGTTGGCCCGGTACAGACCGGCTGCTGA 965
QY 484 CTCCCATCCCTCTGGAT 500

Db 966 CTTCCATCCTTGAT 982

RESULT 9

AAS31002 standard; cDNA; 1393 BP.

AAS31002;

04-DEC-2001 (first entry)

Human diagnostic and therapeutic polynucleotide (DITHP) #17.

Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;

cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;

acquired immune deficiency syndrome; AIDS; autoimmune disorder;

respiratory disorder; ss.

Homo sapiens.

WO200162927-A2.

30-AUG-2001.

21-FEB-2001; 2001WO-US06059.

24-FEB-2000; 2000US-0184693.

24-FEB-2000; 2000US-0184697.

24-FEB-2000; 2000US-0184598.

24-FEB-2000; 2000US-0184768.

24-FEB-2000; 2000US-0184770.

24-FEB-2000; 2000US-0184771.

24-FEB-2000; 2000US-0184772.

24-FEB-2000; 2000US-0184773.

24-FEB-2000; 2000US-0184774.

24-FEB-2000; 2000US-0184776.

24-FEB-2000; 2000US-0184777.

24-FEB-2000; 2000US-0184797.

24-FEB-2000; 2000US-0184813.

24-FEB-2000; 2000US-0184837.

24-FEB-2000; 2000US-0184841.

24-FEB-2000; 2000US-0185213.

12-MAY-2000; 2000US-0203785.

15-MAY-2000; 2000US-0204226.

16-MAY-2000; 2000US-0204525.

16-MAY-2000; 2000US-0204821.

16-MAY-2000; 2000US-0204908.

17-MAY-2000; 2000US-0205332.

17-MAY-2000; 2000US-0204815.

17-MAY-2000; 2000US-0204863.

17-MAY-2000; 2000US-0205221.

17-MAY-2000; 2000US-0205285.

17-MAY-2000; 2000US-0205286.

17-MAY-2000; 2000US-0205287.

17-MAY-2000; 2000US-0205323.

17-MAY-2000; 2000US-0205324.

(INCY-) INCYTE GENOMICS INC.

Chen SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;

Panzer A, D'Sa SA, Amesbury S, Dahl CR, Dam TC, Daniels SE;

Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;

Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockreiter TK, Daffo A;

Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;

Cohen HD, Hodgson DM, Lincoln SE, Jackson S;

WPI; 2001-502867/55.

P-PSDB; AAU19431.

Polynucleotides encoding diagnostic and therapeutic proteins, e.g.

enzymes, hormones and receptors, useful in diagnostics and therapeutics

PT -
XX
PS Claim 1; Page 304; 522pp; English.
XX

The invention relates to polynucleotides (I) encoding diagnostic and therapeutic (DITHP) polypeptides (II), which include e.g. enzymes, and proteins involved in growth and development and receptors. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate DITHP expression. For example, (I) and (II) may be used to treat disorders associated with decreased polypeptide expression by rectifying mutations or deletions in a patient's genome, that affect the activity of the DITHP, by expressing inactive proteins or supplementing the patient's own production of them. (I) and (II) may be used to treat diseases, for example, cell proliferative disorder, Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma, leukaemia, autoimmune disorders, and respiratory disorders. Additionally, (I) may be used to produce the DITHP, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) and (II) also complementarily sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies against DITHPs and in assays to identify modulators of DITHP expression and activity. The anti-DITHP antibodies and antagonists may also be used to down regulate expression and activity. The anti-DITHP antibodies may also be used as diagnostic agents for detecting the presence of DITHPs in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and therapeutic (DITHP) polynucleotides of the invention.

SQ Sequence 1393 BP; 375 A; 352 C; 304 G; 360 T; 2 other;

Query Match 65.4%; Score 326.8; DB 22; Length 1393;

Best Local Similarity 88.7%; Pred. No. 16-103; Mismatches 394; Conservative 0; Mismatches 32; Indels 18; Gaps 3;

66 ACTGAGAGACAGGACTGATGATTTCTAGGCGGACTAGAACTCCTAGCTAGCTG 125

245 ACTGAGAGACAGGACTGATGATTTCTAGGCGGACTAGAACTCCTAGCTAGCTG 304

126 GAGGTGACCACTGCTTAAACAGGGGCTTGCAATTGCTCAGCTGACCAAT 185

305 GAGGTGACCGCTCTTAAACAGGGGCTTGCAATTGCTCAGCTGACCAAT 364

186 C-----AGAGACTGACTTAAATGCTAATTAGCAAAAGCAGAGGTAAGAAAT 236

365 CAGGTAGTAAGAGACTGCTAATAATGCTAATTAGCAAAAGCAGAGGTAAGAAAT 424

237 GCCAATCATCTATTGCTGAGACAGAGAGGAGCAACATCGGATATTAACCCAG 296

425 GCCAATCATCTATTGCTGAGACAGAGAGGAGCAACATCGGATATTAACCCAG 484

297 GCATTGAGCTGGCAACAGAGCCCTTTGGGTCCTTCTTGTATGAGAGCTGTT 356

485 GCATTGAGCTGGCAACAGAGCCCTTTGGGTCCTTCTTGTATGAGAGCTGTT 543

357 TTGATGCTATTGACCTGCTTAAATCTGCACTGACCTTCTGTCATGTTCTTAC 416

544 T-----TTTCACTCTATTAATCTTGAACCTGCACTTCTGCTGTTGTTTAC 595

417 GAGCTGAGCTGAGCTTTTCTCACCCTGACCACTGCTTTTGCACCAACCGAGACTG 476

596 GGTTTGAGCTGAGCTTTTCTCACCCTGACCACTGCTTTTGCACCAACCGAGACTG 655

477 CCGCTGACTCCCATCCCTCTGAT 500

656 CCGCTGACTCCCATCCCTCTGAT 679

RESULT 10

AAS65964 standard; cDNA; 893 BP.

Polynucleotides encoding diagnostic and therapeutic proteins, e.g.

enzymes, hormones and receptors, useful in diagnostics and therapeutics

AC AAS65964;
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #1768.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG01777.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 1768; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electron.c format directly from WFO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 893 BP; 223 A; 264 C; 202 G; 204 T; 0 other;
Query Match 61.4%; Score 306.8; 1B 23; Length 893;
Best Local Similarity 87.2%; Pred. No. 8.1e-97;
Matches 389; Conservative 0; Mismatches 37; Indels 20; Gaps 4;
QY 66 ACTGAGACGAGACTAGTGAATTCCTAGGCGGACTAGATCCCTAGGCTGAGCTGG 125
DB 236 ACTGAGAGCAGAGACTAGTGAATTCCTAGGCGGACTAGATCCCTAGGCTGAGCTGG 295
QY 126 GAAGGTGACGAGCTGACCTTTAAACACGCGGCTTGCACTTAGCTCACACCTGACCAT 185
DB 296 GAAGGTAAACATCATCTTAAAGAGGTGCTTGCACTAGCTCACACCGACCCCAT 355
QY 186 C-----AGAGAGCTCACTAAATGCTAATTAGGCAAGAGCAGAGGAGTAAAGAAATA 236
DB 356 CAGGTAAAGAAAGAGAGCCGCTAAATGCTAATTAGGCAAAAACAGAGAGTAAAGAAATA 415

QY 237 GCCATCATCTATTTGCTGAGACACAGACAGAGGACAAATCGGATATTAACCCAG 296
DB 416 GCCATCATCTATTTGCTGAGACACAGCGGAGGAGCAATGATCAAGATTAACCCAG 475
QY 297 GCATTGAGCTGGACACAGAGAGCCCCCTTTGGGTCCCTTCCCTTTGATAGGAGCTGTT 356
DB 476 GCATTGAGCTGGACACAGAGAGCCCCCTTTGGGTCCCTTCCCTTTGATAGGAGCTG 534
QY 357 TTCATGCTATTTCACTATTAATCTTGCACTG--CACTCTTCTGTCATGTTTCTT 414
DB 535 T-----TTTCACTATTAATCTTGCACTGCACTCTTCTGTTGTTGTT 586
QY 415 ACCGCTGAGCTGAGCTTTTGTCAACCGTCCACCACTGCTGTTTGGCACCAGGAGACC 474
DB 587 ACCGCTGAGCTGAGCTTTTGTCAACCGTCCACCACTGCTGTTTGGCACCAGGAGACC 646
QY 475 TGCCGCTGACTCCCATCCCTCTGGAT 500
DB 647 CATTGCTGACTCCCATCCCTCTGGAT 672

RESULT 11

AAS31000
ID AAS31000 standard; cDNA; 849 BP.AC AAS31000;
XXDT 04-DEC-2001 (first entry)
XXDE Human diagnostic and therapeutic polynucleotide (DIRHP) #15.
XXKW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KW respiratory disorder; ss.XX Homo sapiens.
XXOS Homo sapiens.
XXPN WO200162927-A2.
XXPD 30-AUG-2001.
XXPF 21-FEB-2001; 2001MO-US06059.
XXPR 24-FEB-2000; 2000US-0184693.
XXPR 24-FEB-2000; 2000US-0184697.
XXPR 24-FEB-2000; 2000US-0184698.
XXPR 24-FEB-2000; 2000US-0184768.
XXPR 24-FEB-2000; 2000US-0184769.
XXPR 24-FEB-2000; 2000US-0184770.
XXPR 24-FEB-2000; 2000US-0184771.
XXPR 24-FEB-2000; 2000US-0184772.
XXPR 24-FEB-2000; 2000US-0184773.
XXPR 24-FEB-2000; 2000US-0184774.
XXPR 24-FEB-2000; 2000US-0184776.
XXPR 24-FEB-2000; 2000US-0184777.
XXPR 24-FEB-2000; 2000US-0184797.
XXPR 24-FEB-2000; 2000US-0184813.
XXPR 24-FEB-2000; 2000US-0184817.
XXPR 24-FEB-2000; 2000US-0184841.
XXPR 24-FEB-2000; 2000US-0185213.
XXPR 24-FEB-2000; 2000US-0185216.
XXPR 12-MAY-2000; 2000US-0203785.
XXPR 15-MAY-2000; 2000US-0204226.
XXPR 16-MAY-2000; 2000US-0204525.
XXPR 16-MAY-2000; 2000US-0204821.
XXPR 16-MAY-2000; 2000US-0204908.
XXPR 16-MAY-2000; 2000US-0205232.
XXPR 17-MAY-2000; 2000US-0204815.
XXPR 17-MAY-2000; 2000US-0204863.
XXPR 17-MAY-2000; 2000US-0205221.
XXPR 17-MAY-2000; 2000US-0205285.
XX

CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1478 BP; 389 A; 337 C; 364 G; 383 T; 0 other;
SQ

Query Match 59.1%; Score 295.6; DB 23; Length 1478;
Best Local Similarity 85.7%; Pred. No. 9.2e-93;
Matches 382; Conservative 0; Mismatches 44; Indels 20; Gaps 4;

QY 66 ACTGAGACAGGAGTCTGATTTCTAGGCGCAATTAAGAAATCCCTAGGCTTAGCTGG 125
DB 1473 ATTGAAACAGGACGACCTGATCTCTAGGCTGACTTAATAATCCTTAGGCTTAGCTGG 1414
QY 126 GAAGGTGACACGTCACCTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAAT 185
DB 1413 GAAGGTGACCTGATCCACTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAAT 1354
QY 186 C-----AGAGGCTCACTAAATGCTAATTGCGCAAGACAGGAGTAAAGAAATA 236
DB 1353 CAGGTAGTAAAGAGAGCTCACTAAACGCTCATTTAGGCAAAAACAGGAGTAAAGAAATA 1294
QY 237 GCCATCATCTATTGCTGAGAGCAGACAGAGGAGGACAAACATGGGATTTAAACCCAG 296
DB 1293 GCCATCATCTATTGCTGAGAGCAGACAGGAGGAGGAAATGACCCGGAATTTAAACCCAG 1234
QY 297 GCATTGAGCTGGCAACAGACGCCCCCTTTGGGTCCTTCCCTTTGATGAGAGCTGTT 356
DB 1233 GCATTGAGCTGGCAACGCGC-TACCTCTCTGGGTCATCTCCCTTTGATGAGAGCTGTT 1175
QY 357 TTCATGCTATTTCCTCTATTAAATCTTGCAACTG--TACTTTTGTGCTCATGTTTCTT 414
DB 1174 T-----TTTCACTCTATTAAATCTTGCAACTGCAATCTTCTGGTCCGTGTTTGT 1123
QY 415 ACGGCTGAGCTGAGCTTTTGTCTACCGCTGACGACTCTGTTTGCAACGACGAGACC 474
DB 1122 ATGGCTGAGCTGAGCTTTGCTGACAGTCCACACTCTCTGTTTGCCTGACATGAGACC 1063
QY 475 TGCCGCTGACTCCCATCCCTCTGGAT 500
DB 1062 CCGTACTGACTTCCACCCCTCCGGAT 1037

RESULT 13
AAS88392
ID AAS88392 standard; cDNA; 808 BP.
XX
AC AAS88392;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #24196.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
AC P-PSDB; ABG24205.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 24196; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS84189-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 808 BP; 184 A; 234 C; 189 G; 201 T; 0 other;
Query Match 58.5%; Score 292.4; DB 23; Length 808;
Best Local Similarity 85.2%; Pred. No. 8.6e-92;
Matches 380; Conservative 0; Mismatches 46; Indels 20; Gaps 4;

QY 66 ACTGAGACAGGAGTCTGATTTCTAGGCGCAATTAAGAAATCCCTAGGCTTAGCTGG 125
DB 6 ATTGAAACAGGACGACCTGATGCTCTAGGCTGACTTAAGAAATCCCTAGGCTTAGCTGG 65
QY 126 GAAGGTGACACGTCACCTTTAAACAGGGGCTTGCAACTTAGCTCAACCTGACCAAT 185
DB 66 GAAGGTGACCTGATCCACTTTAAACAGGGGCTTGCAACTTAGCTCAACCTGACCAAT 125
QY 186 C-----AGAGGCTCACTAAATGCTAATTAGGCAAAACAGAGGTAAGAAATA 236
DB 126 CAGGTAGTAAAGAGAGCTCACTAAACGCTCATTTAGCAAAAACAGGAGTAAAGAAATA 185
QY 237 GCCATCATCTATTGCTGAGAGCAGACAGGAGGAGGACAAACATGGGATTTAAACCCAG 296
DB 186 GCCATCATCTATTGCTGAGAGCAGACGAGGAGGAGGAAATGACCGGATTTAAACCCAG 245
QY 297 GCATTGAGCTGGCAACAGACGCCCCCTTTGGGTCCTTCCCTTTGATGAGAGCTGTT 356
DB 246 GCATTGAGCTGGCAACGCGC-TACCTCTCTGGGTCACCTCTTTGATGAGAGCTGTTG 304
QY 357 TTCATGCTATTTCCTCTATTAAATCTTGCAACTG--CACTTTTGTGCTCATGTTTCTT 414
DB 305 T-----TTTCACTCTATTAAATCTTGCAACTGCACTCTTCTGGTCCGTGTTTGT 356
QY 415 ACGGCTGAGCTGAGCTTTTGTCTACCGCTCAACCACTGCTTTTGCCACGACCCGAGACC 474
DB 357 ACGGCTTAGCTGAGCTTTCACTGCGCGCTCCACCACTGCTGTTTGTGCGCGTCAACGACC 416
QY 475 TGCCGCTGACTCCCATCCCTCTGGAT 500
DB 417 CACGCTGACTTCCACCCCTCCGAAT 442

RESULT 14
AAS84189/c
ID AAS84189 standard; cDNA; 1243 BP.
XX
AC AAS84189;
XX

DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #19993.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX MO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG20002.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS
PS Claim 1; SEQ ID No 19993; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1243 BP; 285 A; 334 C; 358 G; 266 T; 0 other;
SQ
Query Match 58.2%; Score 290.8; DB 23; Length 1243;
Best Local Similarity 85.0%; Pred. No. 4e-91;
Matches 379; Conservative 0; Mismatches 47; Indels 20; Gaps 4;

Db 1058 GCCAATCATCTATTGCTGAGACAGCGGAGGAGCAATGACCGGATTTAAACCCAG 999
|
Qy 297 GCATTGAGCTGGCAACAGAGCCCCCTTTGGTCCCTTCTTGTATGAGAGCTGTT 356
|
Db 998 GCATTGAGCTGGCAACAGG-7ACCCCTCTGCTGCTACCTCTTGTATGAGAGCTCTG 940
|
Qy 357 TTCATGCTATTTCACCTCTATTAATCTTGCACTG--CACTCTTGTGATGATTTCTT 414
|
Db 939 T-----TTTCACTCTATTAATCTTGCACTGCACTTTCTGTCGCTGTTGTT 888
|
Qy 415 ACGGCTGAGCTGAGCTTTTGTACCGTCCACCACTGCTGTTGCAACCGAGACC 474
|
Db 887 AGGCTGAGCTGAGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 828
|
Qy 475 TGGCGCTGATCCCACTCCCTCTGAT 500
|
Db 827 CGCAGCTGACTTCTACCCCTCCGAT 802
|
RESULT 15
AAK87491
ID AAK87491 standard; DNA; 2629 BP.
XX
XX AAK87491;
AC
XX
XX 07-NOV-2001 (first entry)
DT
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42303.
DE
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
XX
XX MO200157182-A2.
PN
XX
PD 09-AUG-2001.
XX
XX
PF 17-JUN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220363.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230457.
PR 06-SEP-2000; 2000US-0230458.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
XX Disclosure; SEQ ID NO 42303; 3071pp + Sequence listing; English.
XX
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC activity, and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 2629 BP; 763 A; 616 C; 587 G; 663 T; 0 other;

Query Match 57.5%; Score 287.6; DB 22; Length 2629;
Best Local Similarity 84.5%; Pred. No. 8,3e-90;
Matches 377; Conservative 0; Mismatches 49; Indels 20; Gaps 4;

Oy 66 ACTGAGACAGAGACTACTGATTTCTTCTAGCCGACTAAGATTCCTTAAGCTGCG 125
 |||||
Db 81 ACTGAGACAGAGACTACTGATTTCTTCTAGCGCTGCTAAGATTCCTTAAGCTGCG 140
 |||||
Oy 126 GAAGTGACACGCTCCACTTTAAACACGGGGCTTGCAACTTAGCTCACACTGACCAAT 185


```

Db 141 GAAATTGACCACTGCCCTTTAAACAGGGGCTTGCAATTAGCTCACACCCGACCAAT 200
QY 186 C-----AGAGAGCTCACTAAATGCTAATTAGGCAAGACAGGAGGTAAGAAATA 236
Db 201 CAGGTAGTAAAGAGAGCTCACTAAATGCTAATTAGGAAACAGAGGTAAGAAATA 260
QY 237 GCCAATCATCTATTGCTGAGAGCAGACAGAGGGAACACATCGGATATAAACCCAG 296
Db 261 GCCAATCATCTATTGCTGAGAGCAGACAGAGGGAACATGATCAGGATATAAACCCAG 320
QY 297 GCATTGAGCTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTTGATGGAGCTGT 356
Db 321 GCATTGAGCTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTTGATGGAGCTGT 379
QY 357 TTCATGCTATTTCACCTAATTAAATCTTGAACCTG--CACTCTGCTGTCAGTTCCTT 414
Db 380 T-----TTTCACTCTAATTAAATCTTGAACCTGACACTTCTGTGTAAGTGTGTC 431
QY 415 ACGGCTGAGCTGAGCTTTTGTCTACCGTCCACCACTGCTGTGGCCACACCGCAGACC 474
Db 432 ACGGCTGAGCTGAGCTTTTGTCTACCGTCCACCACTGCTGTGGCCGCTGTCAAGACC 491
QY 475 TGGGCTGACTCCCATCCCTCTGAT 500
Db 492 CACAGCTGACTTCATCCCTCTGAT 517

```

Search completed: April 19, 2003, 12:17:39
 Job time : 197.76 secs

Qy 364 TATTCTACTTATTAATCTTGCAACTGCA 393
Db 2900 TATTCTACTTATTAATCTTGCAACTGCA 2929

RESULT 2

US-08-686-878A-50
Sequence 50, Application US/08686878A
Patent No. 5708157
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavalie, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Evans, Cheryl
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/686,878A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-686-878A-50

Query Match 40.0%; Score 200; DB 1; Length 279;

Best Local Similarity 88.2%; Pred. No. 4.8e-60;

Matches 232; Conservative 1; Mismatches 27; Indels 3; Gaps 2;

Qy 133 ACCAGTCCACCTTTAAACAGGGGCTTGCACTAGC¹CACACCGACCAATCAGAGAG 192
Db 1 RCCACATCCACCTTTAAACAGGGGCTTGCAANAAAG¹ANACCTTGACCAATCAGAGAG 60
Qy 193 CTCCTAAATGCTAATTAAGCAAGAGAGGTAAC¹AAATAGCAATCTATTGCG 252
Db 61 NTCANTAAATATATATTGCGCAAAACAGAGGTAAG¹AAATAGCAATCTATTGCG 120
Qy 253 CTGAGAGCAGCAGAGAGGGAACAATGGGATATAA¹CCAGGATTCGAGCTGGCAA 312
Db 121 CTGAGAGCAGCAGAGAGGGAACAATGATCGGATATAA¹CCAGTTTNGAGCCGGCAA 180
Qy 313 CAGCAGCCCCCTTTGGGCTCCCTTCTGTATGGAG¹CTGTTTCATGCTATTTC 370
Db 181 CGGCA-ACCCCTTTGGGCTCCCTTCTGTATGGAG¹CTGTTTCATGCTATTTC 239
Qy 371 CTCTATTAAATCTTGCAACTGCA 393

Db 240 NTNTATTAAATNTTGCAACTGCA 262

RESULT 3

US-08-721-489-4
Sequence 4, Application US/08721489
Patent No. 5786465
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavalie, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721,489
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-721-489-4

Query Match 40.0%; Score 200; DB 1; Length 279;

Best Local Similarity 88.2%; Pred. No. 4.8e-60;

Matches 232; Conservative 1; Mismatches 27; Indels 3; Gaps 2;

Qy 133 ACCAGTCCACCTTTAAACAGGGGCTTGCACTAGC¹CACACCGACCAATCAGAGAG 192
Db 1 RCCACATCCACCTTTAAACAGGGGCTTGCAANAAAG¹ATNACCTTGACCAATCAGAGAG 60
Qy 193 CTCCTAAATGCTAATTAAGCAAGAGAGGTAAC¹AAATAGCAATCTATTGCG 252
Db 61 NTCANTAAATATATATTGCGCAAAACAGAGGTAAG¹AAATAGCAATCTATTGCG 120
Qy 253 CTGAGAGCAGCAGAGAGGGAACAATCGGATATAA¹CCAGGATTCGAGCTGGCAA 312
Db 121 CTGAGAGCAGCAGAGAGGGAACAATGATCGGATATAA¹CCAGTTTNGAGCCGGCAA 180
Qy 313 CAGCAGCCCCCTTTGGGCTCCCTTCTGTATGGAG¹CTGTTTCATGCTATTTC 370
Db 181 CGGCA-ACCCCTTTGGGCTCCCTTCTGTATGGAG¹CTGTTTCATGCTATTTC 239
Qy 371 CTCTATTAAATCTTGCAACTGCA 393
Db 240 NTNTATTAAATNTTGCAACTGCA 262

[illegible]

Db 1055 CCT 1053

RESULT 6
US-08-973-273-1/c

; Sequence 1, Application US/08973273
; Patent No. 6140085
; GENERAL INFORMATION:
; APPLICANT: Dean, Caroline
; APPLICANT: MacKnight, Richard C
; APPLICANT: Bancroft, Ian
; APPLICANT: Lister, Clare K
; TITLE OF INVENTION: Genetic Control of Flowering
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye P.C.
; STREET: 1100 No. 6140085th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,273
; FILING DATE: 01-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01332
; FILING DATE: 03-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 951196.9
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ms Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9763 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; STRAIN: Columbia
; US-08-973-273-1

Query Match 6.2%; Score 30.8; DB 3; Length 9763;
Best Local Similarity 63.5%; Pred. No. 3.5;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 176 CCTGACCATGAGAGCCACTAAATGCTAATTAGCCTAAGAGGAGTAAAGAAAT 235
DB 4845 CTTGAGAGAGAGAACTATATCTGATGTAAGGCTAAGAAACAGGAGAAAGAAA 4786
QY 236 AGCCATCATCTAT 249
DB 4785 GGATCAATCATCAT 4772

RESULT 7
US-08-933-750C-90
; Sequence 90, Application US/08933750C

; Patent No. 5932442
; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purni
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SINTNOT01
; CLONE: 2184712
; US-08-933-750C-90

Query Match 6.1%; Score 30.6; DB 2; Length 1078;
Best Local Similarity 58.3%; Pred. No. 1.2;
Matches 74; Conservative 0; Mismatches 49; Indels 4; Gaps 1;

QY 288 TAAACCCAGGATTCGAGCTGCAACAGCAGCCCCCTTTGGGTCCTTCCCTTGTATG 347
DB 941 TCACTCTGAGAGAGAACTCTGCTACAGAAAGAGCCCTTGGGCTCCCTTCTT---TG 996
QY 348 GAGGCTGTTTCATGCTATTTCACTCTATTTAAATCTTGCACTGACCTTCTGTGCAT 407
DB 997 ATAGAGATTATATAGCCCTTGTTCCTCAATAACTGGGCAATGGAATCCTAGTGTCTAT 1056
QY 408 GTTCTT 414
DB 1057 ACTGCT 1063

RESULT 8
US-09-234-613-90
; Sequence 90, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti

APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Puri
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SININOT01
CLONE: 2184712
US-09-234-613-90

Query Match
Best Local Similarity 58.3%; Score 30.6; DB 3; Length 1078;
Matches 74; Conservative 0; Mismatches 49; Indels 4; Gaps 1;

QY 288 TAAACCCAGCATTCCAGCTGGCAAGCAAGCCCTTGGCTCCCTTGTATG 347
DB 941 TCAACTGTGAAGAGATCTTGTACAGAGAGCCCTTGGCTCCCTTGTATG 996
QY 348 GGAAGCTTTTCAATGCTATTTCACTATTAATCTTGAACCTGACCTCTTGGTCCAT 407
DB 997 ATAGCACTTATATGCTTCTTGTCCCAATAAAGCGGACAGATGATCTAGTGTAT 1056
QY 408 GTTTCCT 414
DB 1057 ACTGCCT 1063

RESULT 9
US-09-739-455-3
Sequence 3, Application US/09739455
Patent No. 6413756
GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF

FILE REFERENCE: CL000653
CURRENT APPLICATION NUMBER: US/09/739,455
CURRENT FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 11827
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(11827)
OTHER INFORMATION: n = A,T,C or G
US-09-739-455-3

Query Match
Best Local Similarity 51.1%; Score 30.6; DB 4; Length 11827;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 117 CCTAGCTGGAGAGTGACACGCTCCACTTTTAAACAGGGGCTTGCAACTGACAC 176
DB 10853 CCGAGCTGCTGAGTGGAGATGCTGCTTTGTTGGCGGCTTGTCTTAATGAGTT 10912
QY 177 CTGACCAATCAGAGAGCTACTTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATA 236
DB 10913 CCTCTTAGATTATTATACACTAAAAAAATTAAGTTTGAAGAAATAGAGATA 10972
QY 237 GCCAATCATCTATTCCTG 257
DB 10973 CAGAAACATGATTTACAG 10993

RESULT 10
US-09-741-154-3/C
Sequence 3, Application US/09741154
Patent No. 6437110
GENERAL INFORMATION:
APPLICANT: BEASLEY, Ellen M. et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL001061
CURRENT APPLICATION NUMBER: US/09/741,154
CURRENT FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 16389
TYPE: DNA
ORGANISM: Human
US-09-741-154-3

Query Match
Best Local Similarity 60.0%; Score 29.8; DB 4; Length 16389;
Matches 46; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 74 ACAGAGCTAGCTGATTTCTTAGCCGACCTAAGATCCCTAAGCTAGCGGAGGTGA 133
DB 1196 ACAGACTAATGTGAATGCTGAGTCAATTAAGATTCCTGATCCAGCTGGGAGAGGG 1137
QY 134 CCAGCTCACCTT 146
DB 1136 ATGCTCGAGCTT 1124

RESULT 11
US-08-961-083-199
Sequence 199, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8512
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 2023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-199

Query Match 5.8%; Score 29; D3 3; Length 2023;
Best Local Similarity 67.2%; Pred. No. 6.1;
Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 203 TGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTGAGAGCAC 262
DB 1711 TGATTTCACTTTAAGATAGCAAGCTAGAGTATAGTACTGATCATTTGGCTCAGATGAC 1770

QY 263 A 263
DB 1771 A 1771

RESULT 12
US-08-961-527-71
Sequence 71, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-71

Query Match 5.8%; Score 29; DB 4; Length 32768;
Best Local Similarity 67.2%; Pred. No. 30;
Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 203 TGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTGAGAGCAC 262
DB 7947 TGATTTCACTTTAAGATAGCAAGCTAGAGTATAGTACTGATCATTTGGCTCAGATGAC 8006

QY 263 A 263
DB 8007 A 8007

RESULT 13
US-09-738-894A-3/c
Sequence 3, Application US/09738894A
Patent No. 6331423
GENERAL INFORMATION:
APPLICANT: GUGGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: CL000636
CURRENT APPLICATION NUMBER: US/09/738,894A
CURRENT FILING DATE: 2000-12-18
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 36651
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(36651)
OTHER INFORMATION: n = A,T,C or G
US-09-738-894A-3

Query Match 5.8%; Score 29; DB 4; Length 36651;
Best Local Similarity 63.8%; Pred. No. 32;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 15 TCCTTTCTGGATGAGGGCAAAAGCGCTGAGATACAGCAATTAATCTTGAACATGAGAGA 74
DB 6891 TCCTAATCTGGAACATGAGCGCTGATGCGTGGGGCTAAGCAACCATCTTGCAGACATGGA 6832

QY 75 CAGGACTAG 83
DB 6831 AAAGGCTAG 6823

RESULT 14
US-08-703-809-2
Sequence 2, Application US/08703809
Patent No. 5716808
GENERAL INFORMATION:
APPLICANT: Raymond, Christopher K.
TITLE OF INVENTION: GENETIC ENGINEERING OF PICHIA METHANOLICA


```

NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/703,809
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 96-18
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3386 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-703-809-2

```

```

Query Match
Best Local Similarity 49.3%; Score 28.8; DB 1; Length 3386;
Matches 75; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 307 TGGCAACAGCAGCCCCCTTGGGTCCTTCCTTTGATGAGAGCTGTTTCATGCTAT 366
DB 856 TGGCAACAGCAGTAGTCTATCTGCTTAATTCATCCACTTGGGAACGCTCTCTTAC 915
QY 367 TTCACTATTAATCTTGAACGCACTCTTCTGTCATGTTTCTAGCGCTCGAGCT 426
DB 916 CCCAGATTCTCAAGCTAATATCTGCCCCCTTGTCTATGTCCTTTCCTCGTACAAGCG 975
QY 427 GAGCTTTGCTCAACGCTCAACCACTGCTGTT 458
DB 976 GAGCTTTGCTCCCATCTCTTGTCTTGT 1007

```

```

RESULT 15
US-08-703-808-2
Sequence 2, Application US/08703808
Patent No. 5736383
GENERAL INFORMATION:
APPLICANT: Raymond, Christopher K.
TITLE OF INVENTION: PREPARATION OF PICHIA METHANOLICA AUXOTROPHIC
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/703,808
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 96-17
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3386 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-703-808-2

```

```

Query Match
Best Local Similarity 49.3%; Score 28.8; DB 1; Length 3386;
Matches 75; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 307 TGGCAACAGCAGCCCCCTTGGGTCCTTCCTTTGATGAGAGCTGTTTCATGCTAT 366
DB 856 TGGCAACAGCAGTAGTCTATCTGCTTAATTCATCCACTTGGGAACGCTCTCTTAC 915
QY 367 TTCACTATTAATCTTGAACGCACTCTTCTGTCATGTTTCTAGCGCTCGAGCT 426
DB 916 CCCAGATTCTCAAGCTAATATCTGCCCCCTTGTCTATGTCCTTTCCTCGTACAAGCG 975
QY 427 GAGCTTTGCTCAACGCTCAACCACTGCTGTT 458
DB 976 GAGCTTTGCTCCCATCTCTTGTCTTGT 1007

```

Search completed: April 19, 2003, 14:48:55

Job time : 68.6205 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using bw model

Run on: April 19, 2003, 12:08:18 ; Search time 75.5659 Seconds
(without alignments)
6646.925 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500
Perfect score: 500
Sequence: 1 cccctg99g9g9gctctctt.....tgacctccatccctcgtgat 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	348.2	69.6	1329	7	US-08-979-847-108
2	301	60.2	180557	12	US-10-003-806-6
3	301	60.2	180557	12	US-10-003-806-9
4	284.8	57.0	2946	9	US-10-114-893-134
5	283.8	56.8	2930	10	US-09-902-535-1
6	270.4	54.1	635	7	US-08-979-847-102
7	261.4	52.3	410	10	US-09-880-107-538
8	259.2	51.8	541	10	US-09-864-761-8173
9	232.2	46.4	326014	10	US-09-731-231A-3
10	226.2	45.2	1894	10	US-09-864-761-4444
11	218	43.6	15425	9	US-10-091-504-1654
12	218	43.6	15425	10	US-09-764-869-1654
13	204.6	40.0	569	10	US-09-864-761-14951
14	200	40.0	279	12	US-10-040-916-50
15	126	25.2	246	10	US-09-864-761-20462
16	118.2	23.6	446	10	US-09-811-284-3
17	111.6	22.3	475	10	US-09-864-761-895
18	109.4	21.9	3524	10	US-09-972-724-1
19	102	20.4	409	10	US-09-864-761-4153

C 20	101	20.2	559	10	US-09-864-761-7501	Sequence 7501, Ap
C 21	97.8	19.6	504	10	US-09-864-761-7027	Sequence 7027, Ap
C 22	93	18.6	440	10	US-09-864-761-3694	Sequence 3694, Ap
C 23	79.2	15.8	579	10	US-09-864-761-13678	Sequence 13678, A
C 24	50.4	10.1	88	10	US-09-864-761-20907	Sequence 20907, A
C 25	49.8	10.0	525	10	US-09-893-737-31	Sequence 31, Appl
C 26	48.8	9.8	387	10	US-09-864-761-30194	Sequence 30194, A
C 27	37.8	7.6	275	10	US-09-864-761-17675	Sequence 17675, A
C 28	34.6	7.0	1568	9	US-09-735-713A-7	Sequence 7, Appl1
C 29	34.6	6.9	79	10	US-09-864-761-31488	Sequence 31488, A
C 30	32.6	6.5	384	10	US-09-783-590-10544	Sequence 10544, A
C 31	32.6	6.5	653	9	US-10-184-644-402	Sequence 402, App
C 32	32.6	6.5	653	9	US-10-184-634-402	Sequence 402, App
C 33	31.8	6.4	468	9	US-09-796-692-4666	Sequence 4666, Ap
C 34	31.2	6.2	473	10	US-09-864-761-11001	Sequence 11001, A
C 35	31.2	6.2	10514	10	US-09-764-877-3470	Sequence 3470, Ap
C 36	31.2	6.2	197997	10	US-09-822-246-3	Sequence 3, Appl1
C 37	31	6.2	407	10	US-09-864-761-20523	Sequence 20523, A
C 38	30.8	6.2	2000	9	US-09-938-842A-2796	Sequence 2796, Ap
C 39	30.6	6.1	410	10	US-09-867-701-4129	Sequence 4129, Ap
C 40	30.6	6.1	1078	10	US-09-840-787-90	Sequence 90, Appl
C 41	30.6	6.1	8220	10	US-09-797-908-3	Sequence 3, Appl1
C 42	30.6	6.1	8522	10	US-09-817-181-3	Sequence 3, Appl1
C 43	30.2	6.0	414	9	US-10-123-153-418	Sequence 418, App
C 44	30.2	6.0	2012	9	US-09-746-783-147	Sequence 147, App
C 45	30.2	6.0	2014	9	US-10-036-041-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-08-979-847-108
; Sequence 108, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TURE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: MPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 108:

[illegible]

```
QY 266 AGGAGGACAACTCGGGATATTAACCCAGGACTTCGAGCTGGCAACAGAGCCCTT 325
| | | | |
Db 58782 AGGAGGACAACTCGGGATATTAACCCAGGACTTCGAGCTGGCAACAGAGCCCTT 58724
QY 326 TTGGGTCCCTCCCTTTGATGGAGAGCTGTTTATGCTATTTTCACTTATTAATCTTG 385
| | | | |
Db 58723 TTGGGTCCCTCCCTTTGATGGAGAGCTGTTTATGCTATTTTCACTTATTAATCTTG 58672
QY 386 CAACTGCA--CTCTTGTGTCATGTTTCTTACGAGCTCGAGCTGTTTGTCTACCGT 443
| | | | |
Db 58671 CAGCTGCACTCTCTTTGTGTATCTTGTATGTTGTGAGTTGAGCTTGTCTCTGCGCT 58612
QY 444 CCACCACTGCTGTTTGGCCACCGCAGACCTGCGCTGACTCCCATCCCTGAGT 500
| | | | |
Db 58611 CCACCACTGCTGTTTGGCCGCTGTGCGAGACCTGCTGCTGATTCATCCGTCAGAT 58555

RESULT 4
US-10-114-893-134
; Sequence 134, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallee, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Carlin-Duckett, McKeough
; APPLICANT: Kelleher, Kerry S.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6000-10A
; CURRENT APPLICATION NUMBER: US/10/114,893
; EARLIER FILING DATE: 2002-04-02
; EARLIER FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 134
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-114-893-134

Query Match 57.0%; Score 284.8; DB 9; Length 2946;
Best Local Similarity 95.5%; Pred. No. 2.6e-89;
Matches 315; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

QY 66 ACTAGAGACAGACTAGCTGATTTCTTAGCCGACTAAGATCCCTAGCTGG 125
| | | | |
Db 2601 ACTAGAGACAGACTAGCTGATTTCTTAGCCGACTAAGATCCCTAGCTGG 2660
QY 126 GAAGGTACACGCTCCACTTAAACAGGGGCTTGCAACTAGCTACCTGCAAT 185
| | | | |
Db 2661 GAAGGTACACATCCACTTAAACAGGGGCTTGCAACTAGCTACCTGCAAT 2720
QY 186 CAGAGGCTCACTAAATGCTAATTAGCAAGAAGAGAGGTAAAGAAATAGCAATCAT 245
| | | | |
Db 2721 CAGAGGCTCACTAAATGCTAATTAGCAAGAAGAGAGGTAAAGAAATAGCAATCAT 2780
QY 246 CTATTGCTGAGACAGCAGAGAGGCAACAATCGGATATTAACCCAGGCAATTGAG 305
| | | | |
Db 2781 CTATTGCTGAGACAGCAGAGAGGCAACAATCGGATATTAACCCAGGCAATTGAG 2840
QY 306 CTGGCAACAGAGCCCTTTGGTCCCTTCCCTTGTATGGAG--CTGTTTTCATGC 363
| | | | |
Db 2841 CTGGCAACAGGCA-AACCCCTTTGGTCCCTTCCCTTGTATGGAGCTCTGTTTTCATGC 2899
QY 364 TATTCACTCTATTAAATCTTGCAACTGCA 393
```

```
Db 2900 TATTCACTCTATTAAATCTTGCAACTGCA 2929
| | | | |

RESULT 5
US-09-902-535-1
; Sequence 1, Application US/09902535
; Patent No. US20020102530A1
; GENERAL INFORMATION:
; APPLICANT: Keith, Jr., James C.
; APPLICANT: McCoy, John M.
; APPLICANT: Mi, Sha
; TITLE OF INVENTION: Methods and compositions for diagnosing
; TITLE OF INVENTION: and treating preclampsia and gestational trophoblast
; FILE REFERENCE: GIN-6006B4
; CURRENT APPLICATION NUMBER: US/09/902,535
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,657
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (930)...(2546)
US-09-902-535-1

Query Match 56.8%; Score 283.8; DB 10; Length 2930;
Best Local Similarity 95.4%; Pred. No. 5.8e-89;
Matches 314; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

QY 66 ACTAGAGACAGACTAGCTGATTTCTTAGCCGACTAAGATCCCTAGCTGG 125
| | | | |
Db 2603 ACTAGAGACAGACTAGCTGATTTCTTAGCCGACTAAGATCCCTAGCTGG 2662
QY 126 GAAGGTACACGCTCCACTTAAACAGGGGCTTGCAACTAGCTACCTGCAAT 185
| | | | |
Db 2663 GAAGGTACACATCCACTTAAACAGGGGCTTGCAACTAGCTACCTGCAAT 2722
QY 186 CAGAGGCTCACTAAATGCTAATTAGCAAGAAGAGGTAAAGAAATAGCAATCAT 245
| | | | |
Db 2723 CAGAGGCTCACTAAATGCTAATTAGCAAGAAGAGGTAAAGAAATAGCAATCAT 2782
QY 246 CTATTGCTGAGACAGCAGAGAGGCAACAATCGGATATTAACCCAGGCAATTGAG 305
| | | | |
Db 2783 CTATTGCTGAGACAGCAGAGAGGCAACAATCGGATATTAACCCAGGCAATTGAG 2842
QY 306 CTGGCAACAGAGCCCTTTGGTCCCTTCCCTTGTATGGAG--CTGTTTTCATGC 363
| | | | |
Db 2843 CTGGCAACAGGCA-AACCCCTTTGGTCCCTTCCCTTGTATGGAGCTCTGTTTTCATGC 2901
QY 364 TATTCACTCTATTAAATCTTGCAACTGCA 392
| | | | |
Db 2902 TATTCACTCTATTAAATCTTGCAACTGCA 2930

RESULT 6
US-08-979-847-102
; Sequence 102, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
```

APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 538
LENGTH: 410
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA250958
US-09-880-107-538

Query Match 52.3%; Score 261.4; DB 10; Length 410;
Best Local Similarity 91.2%; Pred. No. 1,4e-81;
Matches 300; Conservative 0; Mismatches 26; Indels 3; Gaps 2;

QY 66 ACTAGAGACAGAGACTACTGATTCCTAGGCGCAGCTAAGATCCCTAAGCCTAGCTGG 125
DB 329 ACTAGAGACAGAGACTACTGATTCCTAGGCGCAGCTAAGATCCCTAAGCCTAGCTGG 270
QY 126 GAAAGTGAACCAAGTCCACTTTAAACACGAGGCTTGCACTTAGCTCAGACCTGACCAAT 185
DB 269 GAAAGTGAACCAAGTCCACTTTAAACACGAGGCTTGCACTTAGCTCAGACCTGACCAAT 210
QY 186 CAGAGAGCTCTAATAATGCTAATTAGGCAAGACAGAGAGGTAAAGAAATAGCCATCAT 245
DB 209 CAGAGAGCTCTAATAATGCTAATTAGGCAAGAAACAGAGAGGTAAAGAAATAGCCATCAT 150
QY 246 CTATTGCGCTGA-GAGCAAGACAGAGAGGAGCAACAATCGGGATTAATACCAGGCAATTCGA 304
DB 149 CTATTGCGCTGAAGACACAGTGGAGGAGCAAGATTGCAATTAATACCAGGCAATTCGA 90
QY 305 GCTGGAACACAGAGCGCCCTTTGGGTCCTTCCCTTTGTATGGAG--CTGTTTTCATG 362
DB 89 GCGAGCGAAGGCAACCGCCCTTTGGGTCCTTCCCTTTGTATGGAGGCTGTTTTCACT 30
QY 363 CTATTCACTCTATTAAATCTTGCACTG 391
DB 29 CTATTCACTCTATTAAATCTTGCACTG 1

RESULT 8
US-09-864-761-8173/c
Sequence 8173, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27

```

; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 8173
; LENGTH: 541
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC016663.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; US-09-864-761-8173

Query Match          51.8%; Score 259.2; DB 10; Length 541;
Best Local Similarity 81.6%; Pred. No. 1e-80;
Matches 367; Conservative 0; Mismatches 63; Indels 20; Gaps 5;

```

```

; DB 186 ATTCGGCTTGAATGAGCTTTTCTCACCATCCACACCTGATGATGCTGTGCTGAG 127
; DB 472 AC-CTGCGCTGAGCTCCCATCCCTGGAT 500
; DB 126 ACACCGCTGCTGACTTCACACCTCCGGAT 97

RESULT 9
US-09-731-231A-3/C
; Sequence 3, Application US/09731231A
; Patent No. US2002082189A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001007
; CURRENT APPLICATION NUMBER: US/09/731,231A
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 326014
; LENGTH: 326014
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(326014)
; OTHER INFORMATION: n = A,T,C or G
; US-09-731-231A-3

Query Match          46.4%; Score 232.2; DB 10; Length 326014;
Best Local Similarity 86.7%; Pred. No. 1.2e-69;
Matches 294; Conservative 0; Mismatches 33; Indels 12; Gaps 3;

```

```

CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine ver. 1.1
SEQ ID NO 4444
LENGTH: 1894
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC002346.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444

Query Match      45.2%; Score 226.2; DB 10; Length 1894;
Best Local Similarity 86.4%; Pred. No. 8.2e-69;
Matches 291; Conservative 0; Mismatches 28; Indels 18; Gaps 3;

```

```

237 GCCAATCATCTATTGCTGTGAGACACAGAGGAGCAACAATCGGATATAAACCAG 236
1687 GCCAATCATCTATTGCTGTGAGACACAGAGGAGCAATGATCGGATATAAACCAG 1746
297 GCATTGAGCTGGCAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 356
1747 GCATTGAGCTGGCAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1805
357 TTGATGCTATTGCTATTGCTATTGCTATTGCTATTGCTATTGCTATTGCTATTGCT 393
1806 T-----CTTCACTCTATTGCTATTGCTATTGCTATTGCTATTGCTATTGCTATTGCT 1834

RESULT 11
US-10-091-504-1654/c
Sequence 1654, Application US/10091504
Publication No. US2003005908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1654
LENGTH: 15425
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1654

Query Match      43.6%; Score 218; DB 9; Length 15425;
Best Local Similarity 85.5%; Pred. No. 2e-65;
Matches 284; Conservative 0; Mismatches 30; Indels 18; Gaps 3;

```


NUMBER OF SEQ ID NOS: 2442
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1654
LENGTH: 15425
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-1654

Query Match 43.6%; Score 218; DB 10; Length 15425;
Best Local Similarity 85.5%; Pred. No. 2e-65;
Matches 284; Conservative 0; Mismatches 30; Indels 18; Gaps 3;

QY 71 GAGACAGAGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTAAGCTTGGGAGG 130
DB 1998 GAGACAGAGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTAAGCTTGGGAGG 1939
QY 131 TGACGACGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTAAGCTTGGGAGG 186
DB 1938 TGACGACGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTAAGCTTGGGAGG 1879
QY 187 -----AGAGAGCTCTAATAATGCTAATTAGGCAAGAGAGTAAAGAAATAGCCAA 241
DB 1878 ATTAAGAGAGCTCTAATAATGCTAATTAGGCAAGAGAGTAAAGAAATAGCCAA 1819
QY 242 TCATCTATTGCTTGAAGCAGCAGAGAGGAGCAACATCGGATATTAACCCAGGCAAT 301
DB 1818 TCATCTATTGCTTGAAGCAGCAGAGAGGAGCAACATCGGATATTAACCCAGGCAAT 1759
QY 302 CGAGCTGGAG 361
DB 1758 TGAGCTGGAG 1704
QY 362 GCTATTTCACTTATTAATCTTGCAACTGCA 393
DB 1703 -----TTTCACTTATTAATCTTGCAACTGCA 1676

RESULT 13
US-09-864-761-14951/c
; Sequence 14951, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 14951
;; LENGTH: 569
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AP000233.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.3
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
US-09-864-761-14951

Query Match 40.9%; Score 204.6; DB 10; Length 569;
Best Local Similarity 79.4%; Pred. No. 1.6e-61;
Matches 336; Conservative 0; Mismatches 64; Indels 23; Gaps 7;

QY 66 ACTGAGAGACAGAGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTAAGCTTGG 125
DB 415 ACCGAGAGACAGAGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTAAGCTTGG 356
QY 126 G-AAGGTGACGAGCTCAGCTTTAAACAGGAGGCTTGCAACTTGTGCACTGACCA 184
DB 355 GAAAGGTGACCAAGCTTAAACAGGAGGCTGTAAGTGTGCACTGACCTGACCA 296
QY 185 TC-----AGAGAGCTCACTTAATGTCTAATTGAGC-AAAGACAGAGAGTAAAGAA 234
DB 295 TCAGGTAGTAAAGAGGCTTCACTGAAATCAATTTGAGCTAAAGAGAGGAGTAAAGAA 236
QY 235 TAG-CAATCATCTATTGCTGAGAGCAGAGAGAGAGAGCAATCGGATAT-AAAC 292
DB 235 TAGTCAATCATCTATTGCTGAGAGCAGAGAGAGAGAGCAATCGGATAT-AAAC 176
QY 293 CCAGGAGTTCAGAGCTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 352
DB 175 CCAGGAGTTCAG 116
QY 353 TGTTCATGCTATTCTGCTATTAAATCTTGCAACTG--CACTTTTGTGCTCATGTT 410
DB 115 TCTGT-----TTTCACTCTGTTAAATCTTGCAACTGCTGCTGCTGCTGCTGCT 64
QY 411 TCTTACGAGCTGAGAGCTTGTGCTACCCGTCACCACTGCTGTTTGGCAGCAGCA 470
DB 63 TGTTCGAGCTCAAGCTGAACTTGTGCTACCCGTCACCACTGCTGTTTGGCAGCAGCA 4
QY 471 GAC 473
DB 3 GAC 1

RESULT 14
US-10-040-916-50
; Sequence 50, Application US/10040916
; Patent No. US2002014679A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John
; Lavallie, Edward

Racie, Lisa
Meiberg, David
Treacy, Maurice
Evans, Cheryl
Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040,916
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/887,029
FILING DATE: 07-FEB-1997
APPLICATION NUMBER: 08/686,878
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-040-916-50
Query Match 40.0%; Score 200; D3 12; Length 279;
Best Local Similarity 88.2%; Pred. No. 4.6e-60;
Matches 232; Conservative 1; Mismatches 27; Indels 3; Gaps 2;
QY 133 ACCAGGTCCACCTTTAAACACGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAG 192
DB 1 RCACATCCACCTTTAAACACGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAG 60
QY 193 CTCCTAAATGCTAATTAGGCAAGAGAGGTTAAAGAAATAGCCATCATCTATTGC 252
DB 61 NTCANTAAATGATNATTTGGCAAAAACAGAGGTTAAAGAAATAGCCATCATCTATTGC 120
QY 253 CTGAAGACACACAGAGAGGAGCAACAATGGGATATAACCCAGGATTCGGCTGGCAA 312
DB 121 CTGAAGACACACAGAGAGGAGCAACAATGGGATATAACCCAGGATTCGGCTGGCAA 180
QY 313 CAGCAGCCCCCTTGGGTCCTTCCCTTTGTATGGAGAC--TGTTTCATGCTATTTC 370
DB 181 CGGCA-ACCCCTTTGGGTCCTTCCCTTTGTATGGAGACCTTGTATGCTATTTC 239
QY 371 CTCCTAAATGCTTGGCACTGCA 393
DB 240 TMTATTAAATTTGCACTGCA 262

RESULT 15
US-09-864-761-20462/c
; Sequence 20462, Application US/09864761

Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 20462
LENGTH: 246
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010951.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
OTHER INFORMATION: NT HIT: AB026898.1, EVALUATE 4.00e-57
OTHER INFORMATION: EST_HUMAN HIT: A149205.1, EVALUATE 3.00e-49
OTHER INFORMATION: SWISSPROT HIT: Q02279, EVALUATE 7.30e-01
US-09-864-761-20462
Query Match 25.2%; Score 126; DB 10; Length 246;
Best Local Similarity 85.6%; Pred. No. 4.5e-34;
Matches 178; Conservative 0; Mismatches 20; Indels 10; Gaps 3;

QY 246 CTATTGCTGAGACACAGCAGAGGACAAATCGGATATTAACCCAGCATTCGAG 305
|||
Db 246 CTGTTGCTGAGACACAGCGGAGGACATTAATCAAGATATAACCCAGCATTCGAG 187
|||
QY 306 CTGGCAACAGAGCCCCCTTTGGGTCCCTTTGTTATGGAGCTGTTTCATGCTA 365
|||
Db 186 CTGGCAAGGTA-ACCCCCTTTGGGTCCCTTTGTTATGGAGCTCTAT----- 136
|||
QY 366 TTCACTCTATTAAATCTTGCACTGCACTCTCTGGTGCATGTTCTTAAGGCTCGAGC 425
|||
Db 135 CTTACTCTATTAAATCTTGCACTGTACTCTTCTGGTCCGTTGTAC-GCTTAGAC 77
|||
QY 426 TGAAGTTTGGCTCACCGTCCACCACTGC 453
|||
Db 76 TGAAGTTTGGCTCGCCATCCACCACTGC 49
|||

Search completed: April 19, 2003, 14:55:18
Job time : 251.566 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 11:45:23 ; Search time 1265.98 Seconds
(without alignments)
6396.429 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Perfect score: 500

Sequence: 1 cccctggggcggcgtctctcctt.....tgactcccatccctcgtgat 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST.*
1: em_estbda.*
2: em_estbda.*
3: em_estln.*
4: em_estln.*
5: em_estln.*
6: em_estln.*
7: em_estln.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estfun.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rnd.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	336	67.2	727	17	AG029908 Pan trogl
2	332.2	64.4	683	17	AG102951 Pan trogl
3	285.4	57.1	440	12	BE732673 BE732673
4	283.8	56.8	415	9	AI128496 AI128496
5	283.2	56.6	436	9	AI128526 AI128526
6	282.2	56.4	494	9	AA781423 AA781423

C 7	280	56.0	342	9	AA860368	AA860368 aj59c05.s
C 8	277.2	55.4	921	17	BH149565	BH149565 ENT048TR
C 9	273	54.6	422	14	NS3177	NS3177 yv56h11.s1
C 10	271.8	54.4	388	14	H01325	H01325 y199e01.s1
C 11	271.2	54.2	385	14	NS5091	NS5091 yv43e03.s1
C 12	270.8	54.2	342	14	NS5091	NS5091 yb10h02.s1
C 13	270.6	54.1	771	13	BI087886	BI087886 602852630
C 14	267.4	53.5	522	10	AW971553	AW971553 EST383642
C 15	265	53.0	363	14	T69704	T69704 yd13a03.s1
C 16	263.6	52.7	328	14	AA729556	AA729556 nx58c05.s
C 17	261.4	52.3	410	9	AA250958	AA250958 z807d10.s
C 18	261	52.2	443	9	AA837267	AA837267 oq26d10.s
C 19	260.4	52.1	681	10	AV722664	AV722664 AV722664
C 20	259.6	51.9	425	9	AI570707	AI570707 tm79g09.x
C 21	259.6	51.9	433	9	AI379210	AI379210 cd01g11.x
C 22	259.6	51.9	490	9	AI598135	AI598135 tm14a10.x
C 23	258.8	51.8	681	10	AV721910	AV721910 AV721910
C 24	258.2	51.6	424	14	R27412	R27412 yh46d11.s1
C 25	257.6	51.5	458	14	R76086	R76086 y171b03.s1
C 26	256.2	51.2	404	14	R27389	R27389 yh46a09.s1
C 27	253.4	50.7	893	14	BQ437925	BQ437925 AGENCOURT
C 28	253.2	50.6	446	9	AI393478	AI393478 tg45g04.x
C 29	250.8	50.2	611	17	AG381711	AG381711 RPTC11-16
C 30	246.6	49.3	332	9	AI797629	AI797629 w82a08.x
C 31	245.4	49.1	701	17	AG126669	AG126669 Pan trogl
C 32	243.8	48.8	438	14	R77278	R77278 y175d06.s1
C 33	243.6	48.7	722	17	AG104643	AG104643 Pan trogl
C 34	243.4	48.7	396	9	AA814939	AA814939 oc07d02.s
C 35	239	47.8	722	17	AG049481	AG049481 Pan trogl
C 36	238.6	47.7	326	14	D29167	D29167 HUNNK203 Hu
C 37	238.6	47.7	463	14	R68685	R68685 y114g06.s1
C 38	237.6	47.5	447	12	BF919416	BF919416 QVO-NT015
C 39	236.8	47.4	683	17	AG134524	AG134524 Pan trogl
C 40	236.8	47.0	330	9	AA774109	AA774109 ac36d03.s
C 41	233.6	46.7	470	9	AI791155	AI791155 ab52e07.x
C 42	233.6	46.7	490	12	BF919425	BF919425 QVO-NT015
C 43	232.6	46.5	440	9	AI003607	AI003607 z199a12.s
C 44	232.6	46.5	471	9	AA709471	AA709471 zE91h06.s
C 45	232.4	46.5	757	17	AG121490	AG121490 Pan trogl

ALIGNMENTS

RESULT 1
LOCUS AG029908 727 bp DNA linear GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-002C04.R, genomic survey sequence.
ACCESSION AG029908
VERSION AG029908.1 GI:16556781
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone lib:PTB Chimpanzee Male
ORGANISM Pan troglodytes
BAC Library clone:PTB-002C04.R.
REFERENCE
1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 727)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suenhiro-chou, Tsukuba, Ibaraki, Japan
(E-mail:chimpanzee@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

ACCESSION A1128526
VERSION A1128526.1 GI:3597040
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 436)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 720 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 428.
Location/Qualifiers
1. 436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:1714147"
/clone_lib="Soares_placenta:8to9weeks_2NBHptoc9w"
/dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Placenta; Vector: pTV33 (pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTV33 vector (pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 95 a 96 c 106 g 139 t
ORIGIN
Query Match 56.6%; Score 283.2; DB 9; Length 436;
Best Local Similarity 95.2%; Pred. No. 7e-12;
Matches 314; Conservative 0; Mismatches 13; Indels 3; Gaps 2;
QY 66 ACTGAGACAGACGACTGATTTCTTGGCCGACAAAGATCCCTAAGCCTAGCTGG 125
DB 351 ACTGAAAGACAGGACTGATGATTTCTTGGCTGACTAAGATCCCTAAGCCTAGCTGG 292
QY 126 GAAGGTGACACGCTCCACTTTAAACACGGGGCTTGCACTTAGCTCACCTGACCAAT 185
DB 291 GAAGGTGACACATCCACTTTAAACACGGGGCTTGCACTTAGCTCACCTGACCAAT 232
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT 245
DB 231 CAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT 172
QY 246 CTAATGCTGAGAGACAGCAGAGGAGCAACAATCGGATTAATACCAGGCAATTCGAG 305
DB 171 CTAATGCTGAGAGACAGCAGAGGAGCAACAATCGGATTAATACCAGGCAATTCGAG 112
QY 306 CTGGCAACAGACGCCCCCTTTGGGTCCCTTCCTTTGATGGAG--CTGTTTTCATGC 363
DB 111 CCGGCAAGGCA-ACCCCTTTGGGTCCCTTCCTTTGATGGAGCTGTGTTTTCATGC 53
QY 364 TATTCTACTTATTAATTTTGCAACTGCA 393
DB 52 TATTCTACTTATTAATTTTGCAACTGCA 23
RESULT 6
AA781423/c AA781423 494 bp mRNA linear EST 31-DEC-1998
LOCUS
DEFINITION aj26c03.s1 Soares_testis_NHT Homo sapiens cDNA clone 1391428 3'

similar to contains PTV7.t1 PTV7 repetitive element ;, mRNA sequence.
ACCESSION AA781423
VERSION AA781423.1 GI:2840754
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.llnl.gov/dbip/image/image.html
Insert Length: 1645 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 475.
Location/Qualifiers
1. 494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1391428"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pTV33-Pac (pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adapters (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTV33 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 108 a 111 c 133 g 141 t
ORIGIN
Query Match 56.4%; Score 282.2; DB 9; Length 494;
Best Local Similarity 94.8%; Pred. No. 1.6e-81;
Matches 313; Conservative 0; Mismatches 14; Indels 3; Gaps 2;
QY 66 ACTGAGACAGACGACTGATTTCTTGGCCGACAAAGATCCCTAAGCCTAGCTGG 125
DB 334 ACTGAGACAGGACTGATGATTTCTTGGCTGACTAAGATCCCTAAGCCTAGCTGG 275
QY 126 GAAGGTGACACGCTCCACTTTAAACACGGGGCTTGCACTTAGCTCACCTGACCAAT 185
DB 274 GAAGGTGACACATCCACTTTAAACACGGGGCTTGCACTTAGCTCACCTGACCAAT 215
QY 246 CTAATGCTGAGAGACAGCAGAGGAGCAACAATCGGATTAATACCAGGCAATTCGAG 305
DB 214 CAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT 155
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT 245
DB 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT 155
QY 246 CTAATGCTGAGAGACAGCAGAGGAGCAACAATCGGATTAATACCAGGCAATTCGAG 305
DB 154 CTAATGCTGAGAGACAGCAGAGGAGCAACAATCGGATTAATACCAGGCAATTCGAG 95
QY 306 CTGGCAACAGACGCCCCCTTTGGGTCCCTTCCTTTGATGGAG--CTGTTTTCATGC 363
DB 94 CCGGCAAGGCA-ACNCCCTTTGGGTCCCTTCCTTTGATGGAGCTGTGTTTTCATGC 36
QY 364 TATTCTACTTATTAATTTTGCAACTGCA 393

Db 35 TATTCACTCTATTAAATCTTGCAACTGCA 6

RESULT 7
LOCUS AA860368/c 342 bp mRNA linear EST 31-DEC-1998
DEFINITION aj59c05.s1 Soares_testis_NHT Homo sapiens CDNA clone IMAGE:1394600
3' similar to contains PRR7.c1 PTRS repetitive element ;, mRNA
sequence.
AA860368
ACCESSION AA860368.1 GI:2954363
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 342)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cga@db-rcemail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at:
www.bio.lnl.gov/bdyp/image/image.html
Insert Length: 1305 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 319.
Location/Qualifiers
1. .342
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394600"
/clone_1lb="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc. and primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAAGTGGAGCGGCCCAATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT773 vector. Library
went through one round of normalization to Cos5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 75 a 78 c 82 g 107 t
ORIGIN

Query Match 56.0%; Score 280; DB 9; Length 342;
Best Local Similarity 94.5%; Pred. No. 7.1e-81;
Matches 312; Conservative 0; Mismatches 15; Indels 3; Gaps 2;

Db 66 ACTGAGAGACGAGACTAGCTGATTTCTAGAGCCGACTAAGAAATCCTAGCCTAGCTGG 125
338 ACTGAGAGACGAGACTAGCTGATTTCTAGAGCTGACTAAGAAATCCTAGCCTAGCTGG 279

Db 126 GAAGGTGACCACTCCACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 185
278 GAAGGTGACCACTCCACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 219

Db 186 CAGAGAGCTCACTAAATGCTAATTAGCAAGAAGAGAGGTAAAGAAATAGCAATCAT 245
218 CAGAGAGCTCACTAAATGCTAATTAGCAAGAAGAGAGGTAAAGAAATAGCAATCAT 159

Db 246 CTATTGCTGAGAGACAGCAGAGAGGACCAATCGGAGATTAATCCAGCATTCGAG 305
158 CTATTGCTGAGAGACAGCAGAGAGGACCAATCGGAGATTAATCCAGCATTCGGA 99

Db 306 CTGGCAACAGCAGCCCCCTTTGGGTCCCTTCTTGATGAGAG--CTGTTTCATGC 363
98 GCGGCACAGGCA-AACCCCTTTGGGTCCCTTCTTGATGAGAGCTCGTTTTCATGC 40

Db 364 TATTCACTCTATTAAATCTTGCAACTGCA 393
39 TATTCACTCTATTAAATCTTGCAACTGCA 10

RESULT 8
LOCUS BH149565/c 921 bp DNA linear GSS 27-AUG-2001
DEFINITION ENTQ648TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
Genomic, DNA sequence.
BH149565
ACCESSION BH149565.1 GI:15310303
VERSION
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 921)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HML:IMSS sheared DNA library (2001)
Unpublished (2001)
JOURNAL Contact: Brendan J Loftus
COMMENT department of Eukaryotic Genomics
The Institute for Genomic Research
7712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HML:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 40
High quality sequence stop: 567.
Location/Qualifiers
1. .921
/organism="Entamoeba histolytica"
/strain="HML:IMSS"
/db_xref="taxon:5759"
/clone_1lb="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1, Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 Kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barelli, Oxford University Press, 1999)."

BASE COUNT 242 a 216 c 226 g 237 t
ORIGIN

Query Match 55.4%; Score 277.2; DB 17; Length 921;
Best Local Similarity 84.5%; Pred. No. 9.1e-80;
Matches 377; Conservative 0; Mismatches 48; Indels 21; Gaps 5;

Db 66 ACTGAGAGACGAGACTAGCTGATTTCTAGAGCCGACTAAGAAATCCTAGCCTAGCTGG 125
651 AGTGAAGAGACGAGACTAGCTGATTTCTAGAGCTAAGAAATCCTAGCCTAGCTGG 592

Db 126 GAAGGTGACCACTCCACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 185
591 GAAGGTGACCTGACCTCACTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 532

QY	186	C-----AAGAGGCTACTAAATATGTATTTCACAAAACAAGAGGTAAAGAATA	236
Db	551	CAGTGAATGAAGAGCTCCTCTAAACCGTTAATTGGCAAAAAAGAGCTTAAGAAATA	472
QY	237	GCCAAATCATCTATTTCCTTAGAGCACACAGCAGAGGAGACAAATCGGATATTAACCAG	296
Db	471	GCCAAATCATCTATTTCCTTAGAGCACACCTGGAGAGGAGAACATAT-AGGATATTAACCCAG	413
QY	297	GCATTTCGAGCTGGCAACAGCAGCCCCCTTTGGGTCCTTCCCTTTTGATGGAGCTGTT	356
Db	412	GCATTTCGAGCTGGCAACAGCAGCCCCCTTTGGGTCCTTCCCTTTTGATGGAGCTGTT	354
QY	357	TTCATGCTATTTCACCTATTAAATCTTSCAACTG...CAGCTTCCTGGCCATGTTTTCTT	414
Db	353	T-----TTTCACTCTATTAAATCTTSCAACTGCACTCTTTCTGGTCAAGTTTTGTT	302
QY	415	ACGGCTCGAGCTGAGCTTTTGCTCAACCGTCCACACACTGCTGTTTGGCACACCGAGACC	474
Db	301	ACGACTCGAGCTGAGCTTTCACTTGCCGCCGCCACACACTGCTGTTTGGCCGATCGAGACC	242
QY	475	TGCGCTGACTCCCATCCTCTTGAT	500
Db	241	CGCCCTGACTTCCACCCCTCCAGAT	216
RESULT_9	N53177/c		
LOCUS	DEFINITION	N53177	422 bp mRNA linear EST 28-JAN-1997
		yv6h11.b1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone	
		IMAGE:246789.3 similar to contains PTF7.t2 PTF7 repetitive element	
		; mRNA sequence.	
ACCESSION	N53177		
VERSION	N53177.1	GI:1194343	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 422)		
AUTHORS	Hillier,L., Lennon,G., Becker,M., Bernaldo,M.F., Chiapelli,B.,		
	Chisoso,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins		
	,M., Hillman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore		
	,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,		
	Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,		
	Underwood,K., Wohlmann,P., Welterich,R., Wilson,R. and Marra,M.		
	Generation and analysis of 280,000 human expressed sequence tags		
	Genome Res. 6 (9), 807-828 (1996)		
TITLE	Genome Res. 6 (9), 807-828 (1996)		
JOURNAL	97044478		
MEDLINE	Contact: Wilson RK		
COMMENT	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel.: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LNL ; contact the		
	IMAGE Consortium (info@image.lnl.gov) for further information.		
	Insert length: 938 Std Error: 0.00		
	Seq primer: m13 -40 forward		
	High quality sequence stop: 301.		
FEATURES	Location/Qualifiers		
Source	1..422		
	/organism="Homo sapiens"		
	/db_xref="GDB:3796035"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:246789"		
	/clone_lib="Soares fetal liver spleen INFLS"		
	/sex="male"		
	/dev_stage="20 week-post conception fetus"		
	/lab_host="DH10B (ampicillin resistant)"		
	/note="Organ: Liver and Spleen; Vector: PTF73D (Pharmacia)		
	with a modified polynucleotide site 1: Pac I, Site 2: Eco RI		
	1st strand cDNA was primed with a Pac I - oligo(dT) primer		
	(5' AACGGAAGAATTAATTAAGAATCTTTTATTTTATTTT 3')		

BASE COUNT	95 a	95 c	111 g	120 t	1 others
ORIGIN					
Query Match	54.6%	Score 273;	DB 14;	Length 422;	
Best Local Similarity	93.3%	Ident. No. 1,66-78;			
Matches 308;	Conservative 0;	Mismatches 16;	Indels 6;	Gaps 2;	
<p>double-stranded cDNA was ligated to Eco RI adaptors (pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo."</p>					
OY	66	ACTGAGAGACAGCACTACTGATTCGATTCCTTAGCCGACATAAGATCCCTAAGCTTAGCTGG	125		
DB	327	ACTGAGAGACAGCACTACTGATTCGATTCCTTAGCCGACATAAGATCCCTAAGCTTAGCTGG	268		
OY	126	GAAGGTGACCAACGTCACCTTTAAACACGGGGCTTTGCACTTAGCTCACACCTGACCAAT	185		
DB	267	GAAGGTGACCAACATCCATCTTTAAACACGGGGCTTTGCACTTAGCTCACACCTGACCAAT	208		
OY	186	CAGAGAGCTCACTAATAATGCTAATTAGCCAAAGACAGAGGTAAAGAAATAGCCAAATCAT	245		
DB	207	CAGAGAGCTCACTAATAATGCTAATTAGCCAAAGACAGAGGTAAAGAAATAGCCAAATCAT	148		
OY	246	CTATTGCTCTGAGACAGACAGAGGACCAATCGGGATATATAACCCAGACATTTCAG	305		
DB	147	CTATTGCTCTGAGACAGACAGAGGACCAATCGGGATATATAACCCAGACATTTCAG	88		
OY	306	CTGGCAACAGCAGACGCCCCCTTTGGTCCCTTCCTTTGATATGGAG--CTGTTTTCATGC	363		
DB	87	CGG----CAAGGGCAACCCCTTTGGTCCCTTCCTTTGATATGGAGCTGTGTTTCATGC	32		
OY	364	TATTTCACTCTATTAAATCTTGCAACTGCA	393		
DB	31	TATTTCACTCTATTAAATCTTGCAACTGCA	2		
RESULT 10					
H01325/c		388 bp	mRNA	linear	EST 19-JUN-1995
LOCUS					
DEFINITION		y199e01.61 Soares placenta N28HP	Homo sapiens	cDNA clone	
IMAGE:147384		3', mRNA sequence.			
ACCESSION		H01325			
VERSION		H01325.1	GI:864258		
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE		1 (bases 1 to 388)			
JOURNAL		Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman			
COMMENT		'M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,			
		Ricklin, L., Rohlfing, T., Soares, M., Tan, F., Trevaekis, E., Waterston			
		, R., Williamson, A., Wohlmann, P. and Wilson, R.			
		The Washu-Merck EST project			
		Unpublished (1995)			
		Contact: Wilson RK			
		Washington University School of Medicine			
		4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			
		Tel: 314 286 1800			
		Fax: 314 286 1810			
		Email: ebc@watson.wustl.edu			
		Insert Size: 790			
		High quality sequence stops: 146			
		Source: IMAGE Consortium, LNLN			
		This clone is available royalty-free through LNLN; contact the			
		IMAGE Consortium (info@image.lnl.gov) for further information.			
		Insert Length: 790			
		Seq primer: Promega -21m13			
		High quality sequence stop: 346.			
FEATURES		Location/Qualifiers			
source		1..388			
		/organism="Homo sapiens"			

```

/db_xref="GDB:559031"
/db_xref="taxon:9606"
/clone="IMAGE:147384"
/clone_lib="Soares placenta Nb2Hp"
/sex="Female"
/dep_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pYT3D (pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AATCGAAGAATTGCGCGCCGACAGAAATTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pYT3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bernaldo. "
BASE COUNT      90 a      90 c      96 g      110 t      2 others
ORIGIN

```

Query Match	54.4%	Score 271.8	DB 14	Length 388
Best Local Similarity	93.8%	Pred. No. 3.8e-78		
Matches 305, Conservative	0	Mismatches 17,	Indels 3,	Gaps 2,

QY	66	ACTGAGACACAGGACTTACTGGATTTCCTAGGCGCAGCTAAAGATCCCTAACCTAGCTGG	122
Db	329	ACTGAGACACAGGACTTACTGGATTTCCTAGGCTGACTTAAGATCCCTAACCTAGCTGG	270
QY	126	GAAAGTGAACCACTGTCACCTTTAAACACGGGGCTTGCAACTTACTGCACCTGACCAAT	185
Db	269	GAAAGTGAACCACTGTCACCTTTAAACACGGGGCTTGCAACTTACTGCACCTGACCAAT	211
QY	186	CAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGGAGTTAAAGAAATAGCCAAATCAT	245
Db	209	CAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGGAGTTAAAGAAATAGCCAAATCAT	150

[illegible]

Qy	364	TATTTCACTCTATTAATCTTGCA	388
Db	30	TATTTCACTCTATTAATCCTTGAA	6

RESULT 11	
LOCUS	N55091/c
DEFINITION	N55091 385 bp mRNA linear EST 28-JAN-1997
	yv33603.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
	IMAGE:245500 3', mRNA sequence.

ACCESSION	N55091
VERSION	N55091.1
KEYWORDS	GI:1197970
SOURCE	EST.
ORGANISM	human.
	Homo sapiens

REFERENCE
AUTHORS

Eularyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 385)
Hillier, L., Lennon, G., Becker, M., Donald, M.F., Chippelli, B.,

Chisase, S., Dietrich, N., Dubuque, T., Favello, A., Gish, M., Hawkins, M., Hultman, M., Kucaba, T., Lacey, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Riffitt, L., Rohlfing, T., Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Woldmann, P., Weststrom, R., Wilson, R. and Marra, W. Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)

MEDLINE 97044478
COMMENT Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 266 1800
Fax: 314 266 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LINTL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 907 Std Error: 0.00
Seq primer: m13 -40 forward
High quality sequence crop: 276.
location/Qualifiers
1..385

```

/organism="Homo sapiens"
/db_xref="GDB:3794746"
/cb_xref="taxon:9606"
/clone="IMAGE:245500"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lib_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pTT73D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5', AACTGGAAGATTAATTAAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bonaldo."

```

Query Match	54.2%	Score 271.2	DB 14	Length 385
Best Local Similarity	93.0%	Pred. No. 5.9e-78		
Matches 306	Conservative	0	Mismatches 19	Indels 4
				Gaps 2

[illegible]

Qy	186	CAGAGAGCTCACTAAATATGCTAATTAGCAAGA	CAGAGGTAAAGAAATGGCAATCAT	245
Db	207	CAGAGAGTCTCTAAATATGCTAATTAGCAAGA	CAGAGGTAAAGAAATGGCAATCAT	144
Qy	246	CTATTGCTTGAGACACAGCAGAGGGACAA	CAATCGGATATTAACCCAGCAATTCGAG	305
Db	147	CTATTGCTTGAGACACAGCAGAGGGACAT	GTATGGATATTAACCCAGCAATTCGAG	88

Dy 306 CTGGCAACAGAGCCCCCTTTGGGTCCTTCCCTTGATGGAG--CTGTATTCAATGC 367
| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87 C--GCCACGCGCACCCCCTTTGGGTCCCTCCTTTGTATGGGACTCTGTTTCAATGC 30

QY 364 TATTCTACTATTAATCTGCACTGC 3
|||||
Db 29 TATTCTACTATTAATCTGCACTGC 1

RESULT 12.	
T47345/c	342 bp mRNA linear EST 01-FEB-1995
LOCUS	
DEFINITION	yb10h02.s1 Stratagene placenta (#937225) Homo sapiens cDNA clone
IMAGE:70803	3', mRNA sequence.

ACCESSION	T47345	GI:649326
VERSION	T47345.1	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 342)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M. F., Chiapelli, B.,

Chiesse, S., Dietrich, N., Dubnugue, T., Pavello, A., Gish, M., Hawkins, M., Hiltman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rittfin, L., Rohlfing, T., Schellekens, K., Soares, M.B., Tate, F., Thiermy-Weg, J., Trevaskis, E., Underwood, K., Woldmann, P., Waterston, R., Wilson, R., and Marra, M. Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* 6 (9), 807-828 (1996).

COMMENT Contact: Wilson RK

COMMENT Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 598 Std Error: 0.0)

High quality sequence stop: 343.

FEATURES

Location/Qualifiers

Source

```

/organism="Homo sapiens"
/db_xref="GDB:491700"
/db_xref="taxon:9606"
/clone="IMAGE:70803"
/clone_lib="Stratagene placenta (#937225)"
/sex="male"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: placenta; Vector: paluescript SK-; Site: 1: EcoRI; Site 2: XhoI; Clone: unidirectionally. Primer: Oligo dt. Caucasian. Average insert size: 1.2 kb; Uni-ZAP XR Vector: ~5' adaptor sequence: 5' GAATTCGGACAG 3' ~3' adaptor sequence: 5' CTCGACATTTTTTTTTTTTTTT 3'"
74 a      80 c      80 g      102 t      6 others
BASE COUNT

```

BASE COUNT	74 a	80 c	80 g	102 t	6 others
ORIGIN					

BASE COUNT	74 a	80 c	80 g	102 t	6 others
ORIGIN					

Query Match	54.28;	Score 270.8;	DB 14;	Length 342;
-------------	--------	--------------	--------	-------------

Matches 304; Conservative 0; Mismatches 23; Indels 3; Gaps 2;
Best Local Similarity 92.1%; Freq. NO. 7.0e-16;

Matches 304; Conservative 0; Mismatches 23; Indels 3; Gaps 2;
Best Local Similarity 92.1%; Freq. NO. 7.0e-16;

Qy	66	ACTGAGACACAGACACTACTGGAATTTCTTAGCCGACATAGAAATCCCTAAGCCTAGCTGG	153
Db	331	ACTGAGACACAGACACTACTGGAATTTCTTAGCTGACATAGAAATCCCTAAGCCTAGCTGG	272
Qy	126	GAAAGTGAACACGTCACCTTTTAAACACGGGGCTTGCAACTTACTCAACCTGACCAAT	185
Db	271	GAAAGTGAACACATCCACTTTTAAACACGGGGCTTGCAACTTACTCAACCTGACCAAT	212
Qy	186	CAGAGAGCTCACTAAATAGCTAATTAGGCAAAAGACAGAGTTAAAGAAATAGCCATCAT	245
Db	211	CAGAGAGCTCACTAAATAGCTAATTAGGCAAAAGACAGAGTTAAAGAAATAGCCATCAT	152
Qy	246	CTATTGCTCTGAGAGCACAGCAGAGGGACACAAATCGGAAATTTAAACCCAGGCAATTGGAG	305
Db	151	CTATTGCTCTGAGAGCACAGCAGAGGGACAAATGATCGGAAATTTAAACCCAGGCAATTGGAG	92
Qy	306	CTGGCACAAGAGAGCCCCCTTTGGGTCCCTTCCCTTTGATGGAG--CTGTTTTCAATGC	363
Db	91	GCGGCACAAGGCA-ANCCCTTTGGGTCCCTTCCCTTTGATGGAGCTGTGTTTTCAATGC	33
Qy	364	TATTTCACTCTATTAAATCTTGCAACTGCA	393
Db	32	TATTTCACTCTATTAAANNAANNAATCTGCA	3

RESULT 13						
BI087886	BI087886	771 bp	mRNA	linear	EST 20-JUN-2001	
LOCUS	6028526290P1	NIH_MGC_10	Homo sapiens	cDNA clone IMAGE:4993894	5',	
DEFINITION	mRNA sequence.					
ACCESSION	BI087886					

VERSION	BI087886.1	GI:14506216
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
1 (bases 1 to 771)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.
with NCBI Blast/blast.cgi

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens					
Euryarchaeota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
1 (bases 1 to 522)					
Hedge, P., Qi, R., Abernathy, K., Dharp, S., Gaspard, R., Gay, C., Holt, I. E., Seeded, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and Quackenbush, J.					
Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray unpublished (2000)					
Contact: John Quackenbush					
The Institute for Genomic Research					
9712 Medical Center Dr., Rockville, MD 20850, USA					
Tel: 301 838 3528					
Fax: 301 838 0208					
Email: john@tigr.org					
Plate: 292					
Seq primer: Forward.					
Location/Qualifiers					
1..522					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone_lib="MAGE resequences, MAGU"					
/note="Vector: pbluescriptsm"					
BASE COUNT	125 a	115 c	134 g	148 t	
ORIGIN					
Query Match	53.5%	Score 267.4;	DB 10;	Length 522;	
Best Local Similarity	92.7%;	Pred. No. 1.2e-76;			
Matches 303;	Conservative 0;	Mismatches 21;	Indels 3;	Gaps 2	
QY	66	ACTGAGACACAGAGATAGCTGATTTCTTAGCGCCGACTAAGAACTCCTTAAGCTTACTGG	125		
Db	326	ACTGAGACACAGAGATAGCTGATTTCTTAGCGCCGACTAAGAACTCCTTAAGCTTACTGG	267		
QY	126	GAAGGTGACACAGCTCCACCTTTAAACAGAGGGCTTGCAACTTAAAGCTCAACCTGACCAAT	185		
Db	266	GAAGGTGACACAGCTCCACCTTTAAACAGAGGGCTTGCAACTTAAAGCTCAACCTGACCAAT	207		
QY	186	CAGAGAGCTCACTAAATGCTTAATTAGGCMAAGACAGAGGTTAAAGAAATAGCCATCAT	245		
Db	206	CAGAGAGCTCACTAAATGCTTAATTAGGCMAAGACAGAGGTTAAAGAAATAGCCATCAT	147		
QY	246	CTATTGCTTGAAGACACAGCAGAGGGACACAACTGGGATTTAAACCCAGGCAATTCGAG	305		
Db	146	CTATTGCTTGAAGACACAGCAGAGGGACACAACTGGGATTTAAACCCAGGCAATTCGAG	87		
QY	306	CTGGCAACAGACAGCCCCCTTTGGGATCCCTTCCCTTGTATGGAGG--CTGTTTTCATGC	363		
Db	86	CCAGCAACAGGCA-ACCGGCTTTGGGATCCCTTCCCTTGTATGGAGGCTGTGTTTTCATGC	28		
QY	364	TATTTCACTTAATTAATCTTGCACAT	390		
Db	27	TATTTCACTTAATTAATCTTGCACAT	1		
RESULT 15					
LOCUS	T69704	363 bp	mRNA	linear	EST 07-MAR-1995
DEFINITION	Y133a03.81 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:66988 3', mRNA sequence.				
ACCESSION	T69704				
VERSION	T69704.1	GI:680852			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 363)				
	Hillier, L., Clark, N., Dubuque, T., Ellison, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Woldman, F., and Wilson, R.				

JOURNAL

TITLE The WashU-Merck EST Project
COMMENT Unpublished (1995)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
Insert Size: 766
High quality sequence stops: 341 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 766 Std Error: 0.00
Seq primer: -21m13
High quality sequence stop: 341.

FEATURES

source

```

1..363
    /organism="Homo sapiens"
    /db_xref="GDB:463693"
    /db_xref="taxon:9606"
    /clone="IMAGE:66988"
    /clone_lib="Soares fetal liver spleen INFLS"
    /sex="male"
    /dev_stage="20 week-post conception fetus"
    /lab_host="DH10B (ampicillin resistant)"
    /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' ACTCGAGAGAATTAATTAAAGACTCTTTTCTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      83 a      83 c      87 g     107 t       3 others
ORIGIN
Query Match          53.0%; Score 265; DB 14; Length 363;
Best Local Similarity 89.6%; Pred. No. 6.3e-76;
Matches 294; Conservative 0; Mismatches 33; Indels 1; Gaps 1;
```

QY ACTAGAGACGAGCACTGAGCTGGATTTCCCTAGGCAGACTAAGATCCCTAAGCCTAGCTGG 125
 Db ACTAGAGACGAGCACTGAGCTGGATTTCCCTAGGCAGACTAAGATCCCTAAGCCTAGCTGG 269

QY GAAGGTGACACGTCCTCACTTTAAACAAGGGGCTTGCAACTTAGCTCACACTGAACCAT 185
 Db GAAGGTGACACATCCACTTTAAACAAGGGGCTTGCAACTTAGCTCACACTGAACCAT 209

QY CAGAGAGCTCACTTAAATGCTAATTAGCCAAGACAGAGGTAAAGAAATGCCAATAT 245
 Db CAGAGAGCTCACTTAAATGCTAATTAGCCAAGACAGAGGTAAAGAAATGCCAATCAT 149

QY CTATTGCTGTGAGACACAGCAGAGGGGCAACATCGGGATATATAACCAGACATTCGAG 305
 Db CTATTGCTGTGAGACACAGCAGAGGGGCAATGATCGGGATATATAACCAGATTCGAGAN 89

QY CTGCGAACAGCAGACCCCCCTTTGGGTCCTCTCCCTTTGTATGGAGAGCTGTTTTCAGTGA 365
 Db C-GGGAACGGCAAAACCCCTTTGGGCCCTCCCTTTGTATGGAGAGCTGTTTTCAGTGA 30

QY TTTCACCTCTATTAAATCTTTGCACTGCA 393
 Db TTTCACCTCTATTAAATCTTTNAAAGNACA 2

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:28:30 ; Search time 1700.07 Seconds

(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Perfect score: 500

Sequence: 1 cctctggggcgagctctcctt.....tgactcccatcctcctcgat 500.

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pl:*
10: gb_ro:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_da:*
17: em_fun:*
18: em_fun:*
19: em_in:*
20: em_mu:*
21: em_mu:*
22: em_or:*
23: em_or:*
24: em_or:*
25: em_or:*
26: em_or:*
27: em_or:*
28: em_or:*
29: em_or:*
30: em_or:*
31: em_or:*
32: em_or:*
33: em_or:*
34: em_or:*
35: em_or:*
36: em_or:*
37: em_or:*
38: em_or:*
39: em_or:*
40: em_or:*
41: em_or:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	10499	6 AX007980	AX007980 Sequence
2	500	100.0	56093	6 AX329572	AX329572 Sequence
3	500	100.0	56093	6 HSAC000064	AC000064 Human BAC
4	500	100.0	149194	9 AC007566	AC007566 Homo sapi
5	435	87.0	711	6 AX007997	AX007997 Sequence
6	410	82.0	783	6 AX000970	AX000970 Sequence
7	383	76.8	711	6 AX007998	AX007998 Sequence
8	360	72.0	137482	2 AL161721	AL161721 Homo sapi
9	359	72.0	143590	9 HS83633	AL035706 Human DNA
10	358	71.6	119406	2 AC103595	AC103595 Homo sapi
11	357	71.6	110000	2 AL353584_0	AL353584 Homo sapi
12	357	71.6	169029	9 CNS06C7R	AL390800 Human chr
13	357	71.6	170746	9 AL392173	AL392173 Human DNA
14	357	71.5	163803	9 AC093531	AC093531 Homo sapi
15	357	71.5	191863	9 AC010888	AC010888 Homo sapi
16	356	71.4	180635	9 AL360169	AL360169 Human DNA
17	356	71.3	174019	9 AP001538	AP001538 Homo sapi
18	356	71.3	340000	9 AP001674	AP001674 Homo sapi
19	354	70.8	111079	9 HS419C19	AL035407 Human DNA
20	353	70.7	76169	9 AC003014	AC003014 Human PAC
21	351	70.4	161049	2 AL591842	AL591842 Homo sapi
22	351	70.2	42216	9 AC092899	AC092899 Homo sapi
23	351	70.2	170414	2 AC026019	AC026019 Homo sapi
24	349	69.9	83412	9 AC092843	AC092843 Homo sapi
25	349	69.8	165059	9 AC106856	AC106856 Homo sapi
26	348	69.6	839	9 AF127226	AF127226 Homo sapi
27	348	69.6	1329	6 AX001030	AX001030 Sequence
28	347	69.5	177212	2 AC025733	AC025733 Homo sapi
29	347	69.5	181753	9 AC124075	AC124075 Homo sapi
30	347	69.4	119481	9 AL136234	AL136234 Human DNA
31	347	69.4	152986	9 AC013719	AC013719 Homo sapi
32	346	69.3	179585	2 AC024977	AC024977 Homo sapi
33	346	69.3	185671	9 AC084193	AC084193 Homo sapi
34	346	69.2	156177	9 AC093896	AC093896 Homo sapi
35	345	69.0	37764	9 HSU221F2	275746 Human DNA s
36	345	69.0	143620	9 AC117425	AC117425 Homo sapi
37	344	69.0	157711	9 AC055738	AC055738 Homo sapi
38	344	69.0	181850	9 AC018680	AC018680 Homo sapi
39	344	69.0	207945	9 CENS01DS	AL121841 Human chr
40	344	68.9	318442	2 AC021652	AC021652 Homo sapi
41	344	68.9	140756	9 AL139038	AL139038 Human DNA
42	344	68.9	183499	2 AL607153	AL607153 Homo sapi
43	344	68.9	88328	9 AL357874	AL357874 Human DNA
44	343	68.6	187321	2 AC092510	AC092510 Papio cyn
45	341	68.3	102200	9 HS30M3	AL031775 Human DNA

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AX007980	Sequence 3 from Patent WO9967395.	AX007980	AX007980.1	GI:9995677		human.	Homo sapiens	10499 bp	DNA	linear PAT 06-SEP-2000
AX007980	Sequence 3 from Patent WO9967395.	AX007980	AX007980.1	GI:9995677		human.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 10499)	Perrin,J.P., Rieger,F. and Alliel,P.M.	Nucleic sequence and deduced protein sequence family with human endogenous retroviral motifs, and their uses

JOURNAL Patent: WO 9967395-A 3 29-DEC-1993;
INST NAT SANTE RECH MED (FR); PER:JN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES Location/Qualifiers
Source 1. 10499
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 2495 t
ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 10499;
Best Local Similarity 100.0%; Pred. No. 8.5e-155;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 CCTGTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCCTGGAGATACAGCAATTATC 60
Db 1 CCTGTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCCTGGAGATACAGCAATTATC 60
Qy 61 TTGCAACTGAGAGACAGAGACTAGCTGATTTCTAG;CCGACTAAGAAATCCCTAAGCCTA 120
Db 61 TTGCAACTGAGAGACAGAGACTAGCTGATTTCTAG;CCGACTAAGAAATCCCTAAGCCTA 120
Qy 121 GCTGGGAAGGTGACCAAGCTCCACCTTTAAACAGGGGKCTTGCACTTAGCTCACACTGA 180
Db 121 GCTGGGAAGGTGACCAAGCTCCACCTTTAAACAGGGGKCTTGCACTTAGCTCACACTGA 180
Qy 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAATGACAGAGGTAAGAAATAGCCA 240
Db 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAATGACAGAGGTAAGAAATAGCCA 240
Qy 241 ATCATCTATTGCTGAGAGACAGAGAGGAGCAACAATCGGGATATTAACCCAGGCAAT 300
Db 241 ATCATCTATTGCTGAGAGACAGAGAGGAGCAACAATCGGGATATTAACCCAGGCAAT 300
Qy 301 TCGAGCTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 301 TCGAGCTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Qy 361 TGCTATTTCATCTATTAATCTTGCAACTGACCTCTGTCGATGATGATGATGATGATGAT 420
Db 361 TGCTATTTCATCTATTAATCTTGCAACTGACCTCTGTCGATGATGATGATGATGATGAT 420
Qy 421 CGAGCTGAGCTTTTGTCTACCGCTCCACCACTGCTGTTTGCACACACCGAGAGCTGCCG 480
Db 421 CGAGCTGAGCTTTTGTCTACCGCTCCACCACTGCTGTTTGCACACACCGAGAGCTGCCG 480
Qy 481 TGACTCCCATCCCTCTGGAT 500
Db 481 TGACTCCCATCCCTCTGGAT 500
```

RESULT 2
AX329572 56093 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 81 from Patent WO0194629.
DEFINITION AX329572
ACCESSION AX329572
VERSION AX329572.1 GI:18102550
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ehner, R., Endress, G.,
Hortlgen, S., Soppet, D.R. and Weaver, J.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
Source 1. 56093
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 56093;
Best Local Similarity 100.0%; Pred. No. 9.4e-155;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 CCTGTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCCTGGAGATACAGCAATTATC 60
Db 28001 CCTGTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCCTGGAGATACAGCAATTATC 28060
Qy 61 TTGCAACTGAGAGACAGAGACTAGCTGATTTCTAGGCGCACTAAGAAATCCCTAAGCCTA 120
Db 28061 TTGCAACTGAGAGACAGAGACTAGCTGATTTCTAGGCGCACTAAGAAATCCCTAAGCCTA 120
Qy 121 GCTGGGAAGGTGACCAAGCTCCACCTTTAAACAGGGGCTTGCACTTAGCTCACACTGA 180
Db 28121 GCTGGGAAGGTGACCAAGCTCCACCTTTAAACAGGGGCTTGCACTTAGCTCACACTGA 180
Qy 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 28181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGAGAGAGAGAGAGAGAGAG 28240
Qy 241 ATCATCTATTGCTGAGAGACAGAGAGGAGCAACAATCGGGATATTAACCCAGGCAAT 300
Db 28241 ATCATCTATTGCTGAGAGACAGAGAGGAGCAACAATCGGGATATTAACCCAGGCAAT 300
Qy 301 TCGAGCTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 28301 TCGAGCTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Qy 361 TGCTATTTCATCTATTAATCTTGCAACTGACCTCTGTCGATGATGATGATGATGATGAT 420
Db 28361 TGCTATTTCATCTATTAATCTTGCAACTGACCTCTGTCGATGATGATGATGATGATGAT 420
Qy 421 CGAGCTGAGCTTTTGTCTACCGCTCCACCACTGCTGTTTGCACACACCGAGAGCTGCCG 480
Db 28421 CGAGCTGAGCTTTTGTCTACCGCTCCACCACTGCTGTTTGCACACACCGAGAGCTGCCG 28480
Qy 481 TGACTCCCATCCCTCTGGAT 500
Db 28481 TGACTCCCATCCCTCTGGAT 28500
```

RESULT 3
HSAC000064 56093 bp DNA linear PRI 13-NOV-1996
LOCUS Human BAC clone RG083M05 From 7q21-7q22, complete sequence.
DEFINITION AC000064
ACCESSION AC000064
VERSION AC000064.1 GI:1669369
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 56093)
AUTHORS Paulley, A.
TITLE The sequence of H. sapiens BAC clone RG083M05
JOURNAL Unpublished (1996)
REFERENCE 2 (bases 1 to 56093)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1996)
COMMENT Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
sections once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The
library contains cloned DNA from a human male fibroblast cell line
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBEO

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this
clone is at base position 1 of H_RG083M05; actual end is at 56093
of H_RG083M05

FEATURES

source

This clone contains STS SMS1725.
location/Qualifiers

1..56093

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone_1b="H_RG083M05"

/clone_1b="CITB-978SK-B"

complement(838..1131)

/rpt_family="ALU"

<1360..16971

/gene="WUGSC:H_RG083M05.1"

join(<1360..1503,4181..4370,4587..4774,6422..6556,

9483..9547,11631..11773,11864..12021,13131..13296,

14885..14988,16349..16546,16837..16971)

/gene="WUGSC:H_RG083M05.1"

/note="Atpase; strong similarity to peroxisome

biosynthesis protein PAB1 (PID:g1172019); coded for by

human cDNA C04279 (NID:g1467550)"

/codon_start=1

/protein_id="AAB46346.1"

/db_xref="GI:1669371"

/translation="KRLENIQKTELEVAESEAQMPSVLLDDLDIAGLPAVEHEH

SPDAORCEILCNVYIKKDKDINKFTDPLDQVAKETGEFVARDVLADRIHRL

SRQISTREKLVITLIDROKALRGRLPASRLSNLHKRPGIDKICGAEHVAQIIMD

TIQPAKVCCKKKEYPRLFANLPIRQRTGILVGPRTGTLIAGTARSRNPIISV

KGBELSKYIGASEQAVRDIIRAQAAPCILEFDEFESIAPRGHNTGTVRVNQ

LLTQLDQVEGLQGVYVLAATSRPDLIDPALRGRLDKCYCPDPQVITISLESKTQ

QMHSFVSRLEILNVLNLSLPLADVDLQHVASVDSFGADLKALLVNAQLALHG

MLSKMSEILPDSKEMRYRLVGGSYSESLGNGTSDLSGLSAPSSWTODLPGVP

GKOLPSPQPVLRRTASQGGCELTOEGRDRLADISIKKRYRSQSGEDSMNQPPRI

KTRLAISQSHMTALGHTRPISBDDKNAFEL"

complement(4948..5130)

/rpt_family="ALU"

complement(6581..7133)

/rpt_family="U1"

complement(7767..8037)

/rpt_family="ALU"

complement(8186..8472)

/rpt_family="ALU"

8473..8625

/gene="WUGSC:H_RG083M05.1"

/note="match to human 3' EST H75782 (NID:g1049794), bases

287-444"

8841..9161

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H75921 (NID:g1050050), bases

21-348"

9481..9547

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST N22627 (NID:g1130501), bases

276-343"

repeat_region

complement(12612..12907)

/rpt_family="ALU"

13670..13793

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

143-266"

13794..13877

/rpt_family="ALU"

13878..13906

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

30-58"

13907..14104

/rpt_family="ALU"

complement(14110..14137)

/rpt_family="U1"

complement(15618..15907)

/rpt_family="ALU"

17227..17522

/rpt_family="ALU"

16667..19235

/note="match to human fetal brain 5' EST D61494

(NID:g970409), bases 1-255, and to human 3' EST R07476

(NID:g759399)"

19550..19670

/rpt_family="ALU"

21507..37303

/note="similarity to various SS-RNA virus polyproteins;

pseudogene; region of matches and close matches to

multiple human ESTs, see R68740 (NID:g842257)"

37316..37489

/note="Grail prediction, score = 80"

/evidence=not experimental

complement(138938..39224)

/rpt_family="ALU"

39225..39707

/note="match to multiple human ESTs, see N30113

(NID:g148633)"

39800..40085

/rpt_family="ALU"

complement(40247..40538)

/rpt_family="ALU"

complement(40632..40924)

/rpt_family="ALU"

complement(42283..42891)

/rpt_family="ALU"

complement(45474..45613)

/rpt_family="ALU"

complement(45614..45737)

/note="match to human 3' EST H48898 (NID:g988738), bases

129-333"

complement(46107..47026)

/note="match to multiple human ESTs, see N81064

(NID:g1243765), H48897 (NID:g988737), and M78831

(NID:g273146)"

complement(47027..47318)

/rpt_family="ALU"

complement(47365..47782)

/note="match to multiple human ESTs, see W37495

(NID:g1319089)"

47898..48115

/note="match to human 5' EST H62306 (NID:g1015138), bases

93-368"

complement(48116..48405)

/rpt_family="ALU"

complement(48406..48584)

/note="match to human 3' EST N29952 (NID:g1148472), bases

290-455, and 5' EST R12730 (NID:g765806)"

complement(48787..49405)

/rpt_family="ALU"

complement(49406..49534)

/note="match to human 3' EST R65794 (NID:g838432), bases

309-440"

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBel0BAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 15032 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

source

```
1.149194
Location/Qualifiers
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="7"
  /map="7q21-7q22"
  /clone="CTB-10G5"
  /clone_1lb="CTB-978SK-B"
  1.1634
  /rpt_family="L1"
  2248..2388
  /note="match to EST BG752883 (NID:g14063536)"
  2248..2387
  /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
  2248..2387
  /note="match to EST AW579261 (NID:g7254310)"
  2248..2387
  /note="match to EST BG766882 (NID:g14077535)"
  2248..2387
  /note="match to EST BI160365 (NID:g14620366)"
  2248..2374
  /note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
  2248..2287
  /note="match to EST AV686676 (NID:g10288539)"
  2253..2387
  /note="match to EST AU123510 (NID:g10948226)"
  2344..2387
  /note="match to EST AV686676 (NID:g10288539)"
  2696..3066
  /rpt_family="MAIR"
  3108..3392
  /rpt_family="A1u"
  3540..3628
  /note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
  3540..3628
  /note="similar to Mus musculus EST BB253526 (NID:g8946272)"
  3542..3632
  /note="match to EST AV686676 (NID:g10288539)"
  3542..3628
  /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
  3542..3628
  /note="match to EST AU123510 (NID:g10948226)"
  3542..3628
  /note="match to EST BG766882 (NID:g14077535)"
```

```
misc_feature 3542..3628
              /note="match to EST BI160365 (NID:g14620366)"
misc_feature 3542..3609
              /note="match to EST AW579261 (NID:g7254310)"
misc_feature 3562..3628
              /note="match to EST BE272564 (NID:g9146913)"
misc_feature 3598..3628
              /note="match to EST BE299708 (NID:g9183456)"
misc_feature 3714..3785
              /note="match to EST BG260659 (NID:g12770475)"
misc_feature 3717..3785
              /note="similar to Mus musculus EST BE994936 (NID:g10678674)"
misc_feature 3719..3785
              /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 3719..3785
              /note="match to EST AU123510 (NID:g10948226)"
misc_feature 3719..3785
              /note="match to EST AU123510 (NID:g10948226)"
misc_feature 3719..3785
              /note="match to EST BE272564 (NID:g9146913)"
misc_feature 3719..3785
              /note="match to EST BE299708 (NID:g9183456)"
misc_feature 3719..3785
              /note="match to EST BG766882 (NID:g14077535)"
misc_feature 3719..3785
              /note="match to EST BI160365 (NID:g14620366)"
misc_feature 3719..3785
              /note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature 3860..3862
              /note="match to EST AV686676 (NID:g10288539)"
repeat_region 4730..4827
              /rpt_family="L2"
              4834..4925
              /rpt_family="L1"
              5715..5860
              /note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
              5715..5860
              /note="match to EST AU123510 (NID:g10948226)"
              5715..5860
              /note="match to EST BE272564 (NID:g9146913)"
              5715..5860
              /note="match to EST BE299708 (NID:g9183456)"
              5715..5860
              /note="match to EST BG260659 (NID:g12770475)"
              5715..5860
              /note="match to EST BG766882 (NID:g14077535)"
              5715..5860
              /note="match to EST BI160365 (NID:g14620366)"
              5715..5860
              /note="similar to Mus musculus EST BE994936 (NID:g10678674)"
Query Match 100.0%; Score 500; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 1e-154;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	241	ATCATCTATTGGCTTGAGACACAGCAGAGGGACAC/CAATGGGATATTAAACCCAGGCAT	300
Db	93763	ATCATCTATTGGCTTGAGAGCACAGCAGAGGGACAC/CAATGGGATATTAAACCCAGGCAT	93704
QY	301	TGCAGCTGGCAACAGCAGAGCCCCCTTTGGAGTCCCT/CCCTTGTATGGAGCTGTTTTCA	360
Db	93703	TGCAGCTGGCAACAGCAGAGCCCCCTTTGGAGTCCCT/CCCTTGTATGGAGCTGTTTTCA	93644
QY	361	TGCTATTTTCACTCTATTTAAATTGTGCAACGTGACATCTTGTGTCATATGTTCTTACGGCT	420
Db	93643	TGCTATTTTCACTCTATTTAAATTGTGCAACGTGACAC/TCTGTGTCATATGTTCTTACGGCT	93584
QY	421	CGAGCTGAGCTTTTGTCTCAACGTCACCACTGCTGT TTGCCACCAACCCAGACCTGCCGC	480
Db	93583	CGAGCTGAGCTTTTGTCTCAACCTTCCACCACTGCTGT TTGCCACCAACCCAGACCTGCCGC	93524
QY	481	TGACTTCGCATCCCTCTGGAT	500
Db	93523	TGACTTCGCATCCCTCTGGAT	93504

RESULT		5			PAT 06-SEP-2000
AX007997					
LOCUS	AX007997		711 bp	DNA	linear
DEFINITION	Sequence 20 from Patent WO967395.				
ACCESSION	AX007997				
VERSION	AX007997.1				
KEYWORDS	GI : 9995694				
SOURCE	human.				

CROSS-REFERENCES:
EnuariaVoluta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 711)

REFERENCE
Perin,J.P., Rieger,F. and Alliel,P.⁴,
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
Patent: WO 967395-A 20 28-DEC-1999.

JOURNAL
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR).

FEATURES
Location/Qualifiers

FEATURES		Location/Qualifiers
source	1..711	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
BASE COUNT	170 a 204 c 162 g 175 t	
ORIGIN		
Query Match	87.0%;	Score 435; DE 6; Length 711;
Best Local Similarity	100.0%;	Pred. NO. 3.2e-133;
Matches 435; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

Qy	66	ACGAGAGACAGGA	CTACTGTA	TTTCCCTAGCCGAC	CTAGAAATCCCTAGCCCTACTGG	125		
Db	1	ACTGAGACACAGGA	CTACTGGA	TTTTCCTAGCCGAC	CTAGAAATCCCTAGCCCTACTGG	60		
Qy	126	GAAAGTGAC	CCACGTC	CCACTTTAAA	CACGAGGCTTGC	ACTTAGCTCAC	CTGCACCAAT	185
Db	61	GAAAGTGAC	CCACGTC	CCACTTTAAA	CACGAGGCTTGC	ACTTAGCTCAC	CTGCACCAAT	120
Qy	186	CAGAGAGCTCA	CTAAATGCTAA	TTAGGCANA	GACAGAGGTAAAGAAATAGCCATCAT	245		
Db	121	CAGAGAGCTCA	CTAAATGCTAA	TTAGGCANA	GACAGAGGTAAAGAAATAGCCATCAT	180		
Qy	246	CTATTGCTTGAG	CACAGCAGAGGAGCA	CAATCGG	ATATTAACCCAGGCAATTCGAG	305		
Db	181	CTATTGCTTGAG	CACAGCAGAGGAGCA	CAATCGG	ATATTAACCCAGGCAATTCGAG	240		
Qy	306	CTGGGCAACAG	CAAGCCCCCTTTGGTCCCTTC	CCCTTGGATGGAGAGCTGTTTCATGCA	365			
Db	241	CTGGGCAACAG	CAAGCCCCCTTTGGTCCCTTC	CCCTTGGATGGAGAGCTGTTTCATGCA	300			
Qy	366	TTTCACTATTAA	ATCTTGC	CAACTCTTCTGGT	CATGTTTCTTACGGCTCGAGC	425		
Db	301	TTTCACTATTAA	ATCTTGC	CAACTCTTCTGGT	CATGTTTCTTACGGCTCGAGC	360		

OY	436	TGAGCTTTTGTCTACCGGTCCACCACTCTGTTTGTCCACCAACCGAGACTGCGCGGTACT	485
Db	361	TGAGCTTTTGTCTACCGGTCCACCACTCTGTTTGTCCACCAACCGAGACTGCGCGGTACT	420
OY	486	CCCATCCCTCTGGAT	500
Db	421	CCCATCCCTCTGGAT	435

RESULT 6	AX000970	LOCUS	DEFINITION	ACCESSION	VERSION
	AX000970	783 bp	Sequence 15 from Patent WO902686.	AX000970	AX000970.1
		DNA			GI:7241212
		linear			
					PAT 10-MAR-2000

REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
1 (bases 1 to 783)	Beseme F and Blond J.	ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE DISEASES OR WITH PREGNANT DISORDERS	Patent: WO 9902696-A 15 21-JAN-1999,	
	BIO MERIEUX (FR); BESEME FREDERIC (FR)	Location/Qualifiers		

source	1.	.783	/organism="unidentified"	/db_xref="taxon:32644"
BASE COUNT	173 a	213 c	166 g	180 t
				51 others

Query Match	82.0%	Score 410;	DB 6;	Length 783;
Best Local Similarity	93.8%;	Pred. No. 6.0e-125;		
Matches 408; Conservative	1;	Mismatches	2;	Gaps 1;

[illegible]

RESULT 7			
AX007998			
LOCUS	711 bp	DNA	linear
AX007998			PAT 06-SEP-2000

DEFINITION Sequence 21 from Patent WO967395.
ACCESSION AX007998
VERSION AX007998.1 GI:9995695
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 711)
AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
JOURNAL Patent: WO 967395-A 21 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES
source
1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 171 a 209 c 160 g 171 t
ORIGIN
Query Match 76.8%; Score 383.8; DB 6; Length 711;
Best Local Similarity 95.4%; Pred. No. 3.6e-116;
Matches 417; Conservative 0; Mismatches 17; Indels 3; Gaps 2;
QY 66 ACTGAGAGACGAGCTGCTGATTTCTAGCGGACTAAGATCCCTAAGCTAGCTGG 125
DB 1 ACTGAGAGACGAGCTGCTGATTTCTAGCGGACTAAGATCCCTAAGCTAGCTGG 60
QY 126 GAAAGTGACACGCTGCACTGCTTAAACAGGGGCTTGAACCTAGCTCACCTGACCAAT 185
DB 61 GAAAGTGACACATTCACCTTTAAACAGGGGCTTGAACCTAGCTCACCTGACCAAT 120
QY 186 CAGAGAGCTCACTAATGCTAATTAGGCAAGACAGAGGTAAAGAAATGCCAATCAT 245
DB 121 CAGAGAGCTCACTAATGCTAATTAGGCAAGACAGAGGTAAAGAAATGCCAATCAT 180
QY 246 CTATTTGCTGAGACAGCAGGAGGCAACATCGGGATATTAACCCAGCAATTGAG 305
DB 181 CTATTTGCTGAGACAGCAGGAGGCAACATCGGGATATTAACCCAGCAATTGAG 240
QY 306 CTGCGAAGACAGAGCCCTTTGGGCTCCCTTCCCTTTATAGGAG--CTGTTTTCATGC 363
DB 241 CCGGCAAGAGGA-ACCCCTTTGGGCTCCCTTCCCTTTATAGGAGCTGTTTTCATGC 299
QY 364 TATTTCACTAATTAATCTTGCAACTGCACTCTTGCTCCATGTTTCTTACGGCTCGA 423
DB 300 TATTTCACTAATTAATCTTGCAACTGCACTCTTGCTCCATGTTTCTTACGGCTCGA 359
QY 424 GCTAGAGCTTTGCTCAGCGTCCACACGCTGTTTGCAACGCGGACGACCTGCGGCTGA 483
DB 360 GCTAGAGCTTTGCTCAGCGTCCACACGCTGTTTGCCGACGCGGACGCGCGCTGA 419
QY 484 CTCCATCCCTCTGGAT 500
DB 420 CTCCATCCCTCTGGAT 436
RESULT 8
AL161721/c 137492 bp DNA linear HTG 10-JUL-2001
LOCUS AL161721
DEFINITION Homo sapiens chromosome 6 clone RP1-301F24 map q25.2-26, ***
SEQUENCING IN PROGRESS ***, 6 unordered pieces.
ACCESSION AL161721
VERSION AL161721.3 GI:12330750
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 137492)
AUTHORS Sims,S.

TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: hunquerry@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Jan 22, 2001 this sequence version replaced gi:9212839.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: hunquerry@sanger.ac.uk
----- Project Information
Center project name: DJ301F24
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 135573 bases at least Q40
Consensus quality: 135245 bases at least Q30
Consensus quality: 136166 bases at least Q20
Insert size: 136992; sum-of-contigs
Insert size: 152637; 0.5% error; agarose-fp
Quality coverage: 3.96x in Q20 bases; sum-of-contigs Quality
coverage: 3.56x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 12245: contig of 12245 bp in length
* 12246 12345: gap of 100 bp
* 12346 20428: contig of 8083 bp in length
* 20429 20528: gap of 100 bp
* 20529 85104: contig of 64576 bp in length
* 85105 85204: gap of 100 bp
* 85205 96812: contig of 11608 bp in length
* 96813 96912: gap of 100 bp
* 96913 112622: contig of 15710 bp in length
* 112623 112722: gap of 100 bp
* 112723 137492: contig of 24770 bp in length.
FEATURES
source
1..137492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q25.2-26"
/clone="RP1-301F24"
/clone_11b="RP1-1"
1..12245
/note="assembly_fragment:00933
fragment_chain:1
clone_end:T7
vector_side:left"
12346..20428
/note="assembly_fragment:01071
fragment_chain:1"
20529..85104
/note="assembly_fragment:00093
fragment_chain:1"
85205..96812
/note="assembly_fragment:00954
fragment_chain:1"
96913..112622
/note="assembly_fragment:01341"
112723..137492
/note="assembly_fragment:00790
clone_end:SP6
vector_side:right"
BASE COUNT 37669 a 28026 c 30367 g 40929 t 501 others
ORIGIN

Query Match 72.0%; Score 360; LB 2; Length 137492;
 Best Local Similarity 89.6%; Pred. No. 4.3e-108;
 Matches 398; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

Qy 57 TATCTGCACTGAGACAGAGACTAGCTGATTTCTTAGCCGACTAAGAACTCCCTAAG 116
 Db 128053 TTTATGGTCTTGAGAGACAGACTAGCTGATTTCTTAGCCGACTAAGAACTCCCTAAG 127994
 Qy 117 CCTACTGGGAAGGTGACACGCTCACTTAACAGGGGCTTGCACTTAGCTCACAC 176
 Db 127993 CTTACTTGAAGGTGACGACGACCTTAACAGAGGGCTTGCACTTAGCTCACAC 127934
 Qy 177 CTGACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGTAAAGAAATA 236
 Db 127993 CCACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGTAAAGAAATA 127874
 Qy 237 GCCAATCATCTATTGCTGAGACACAGAGAGGGAACACATGGGATTAACCCAG 296
 Db 127873 GCCAATCATCTCTCTGAGAGACAGAGAGGGAACAGATGGGATTAACCCAG 127814
 Qy 297 GCATTCGAGCTGGCAACAGACCCCTTTGGGTCCCTTCCCTTGTATGGAGCTGT 356
 Db 127813 GCATTCGAGCTGGCAACAGACCCCTTTGGGTCCCTTGTATGGAGCTGT 127755
 Qy 357 TTCATGCTATTTCACCTTAAATCTTGCACTGCATCTTCTGGTCCATGTTCTTAC 416
 Db 127754 TTCATGCTATTTCACCTTAAATCTTGCACTGCATCTTCTGGTCCATGTTCTTAC 127695
 Qy 417 GGCTGAGCTGAGCTTTGCTGACCGCTCCACCATCTGCCTTTTGCAACACCCGACCTG 476
 Db 127694 GGCTGAGCTGAGCTTTGCTGACCGCTCCACCATCTGCCTTTTGCAACACCCG 127635
 Qy 477 CCGCTGACTCCCATCCCTCTGGAT 500
 Db 127634 CCGCTGACTCCCATCCCTCTGGAT 127611

RESULT 9
 HS836J3/C
 LOCUS HS836J3 143590 bp ENA linear PRI 02-MAR-2000
 DEFINITION Human DNA sequence from clone RPS-836J3 on chromosome 1p22.1-31.1
 Contains STSs, GSSs and a putative CpG island, complete sequence.
 ACCESSION AL035706
 VERSION AL035706.10 GI:5777502
 KEYWORDS HMG; CpG island.
 SOURCE Hmo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Crania-ia; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 143590)
 AUTHORS Bagunley C.
 TITLE Direct Submission
 JOURNAL Submitted (29-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humguery@sanger.ac.uk
 REQUESTERS: clonerequest@sanger.ac.uk

COMMENT

This sequence has been finished according to sequence map criteria as follows. An attempt is made to remove all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/chr1
 RPS-836J3 is from the library RPS-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RPS-836J3.

FEATURES

source
 1..143590
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="p22.1-31.1"
 /clone="RPS-836J3"
 /clone_lib="RPS-5"
 1160..1786
 /note="match: GSS: Em: B16380"
 complement(2068..2500)
 /note="match: GSS: Em: A089628"
 2554..3072
 /note="match: GSS: Em: A0118130"
 complement(3824..4133)
 /note="match: GSS: Em: A0089896"
 4017..4402
 /note="match: GSS: Em: A0063212"
 4017..4312
 /note="match: GSS: Em: A005070"
 4023..4278
 /note="match: GSS: Em: A0063209"
 6000..6045
 /note="2 copies 23 mer 100% conserved"
 6037..6906
 /note="ORF repeat: matches 190..467 of consensus"
 9859..10375
 /note="MLTIF repeat: matches 3..536 of consensus"
 13112..13891
 /note="LTR17 repeat: matches 1..780 of consensus"
 15130..15173
 /note="22 copies 2 mer tg 93% conserved"
 15133..15172
 /note="4 copies 10 mer gtgtgtgtgt 95% conserved"
 23083..23118
 /note="18 copies 2 mer tg 80% conserved"
 23328..23713
 /note="WSTD repeat: matches 1..394 of consensus"
 26442..26509
 /note="34 copies 2 mer gt 66% conserved"
 complement(26847..27324)
 /note="match: GSS: Em: A0835456"
 27344..27699
 /note="match: GSS: Em: A0479571"
 31575..32054
 /note="match: STS: Em: AL009625"
 32094..32491
 /note="LTR2D repeat: matches 1..490 of consensus"
 32492..32857
 /note="match: GSS: Em: A0140889"
 32521..32580
 /note="6 copies 10 mer gatagataga 80% conserved"
 32528..32587
 /note="15 copies 4 mer atag 96% conserved"
 32591..32639
 /note="LTR2D repeat: matches 505..553 of consensus"
 38429..38876
 /note="match: GSS: Em: A0146760"
 39727..39756
 /note="6 copies 5 mer tggct 96% conserved"
 40414..40453
 /note="20 copies 2 mer ca 97% conserved"
 42939..42989
 repeat_region

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 119406)
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
COMMENT -----
Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu

Project Information
Center project name: chr-1
Center clone name: RP11-415P10 (sc0362)

Summary Statistics
Sequencing vector: plasmid, L04752; 100% of reads
Chemistry: Dye-terminator ET; 32% of reads
Chemistry: Dye-terminator Big dye; 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 109646 bases at least Q40
Consensus quality: 115473 bases at least Q30
Consensus quality: 117774 bases at least Q20
Insert size: 194800; 40.1% error; agarose-fp
Insert size: 118506; sum-of-contigs
Quality coverage: 5.7x in Q20 bases; agarose-fp
Quality coverage: 9.3x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 4564: contig of 4564 bp in length
* 4565 4664: gap of unknown length
* 4665 12930: contig of 8266 bp in length
* 12931 13030: gap of unknown length
* 13031 20784: contig of 7754 bp in length
* 20785 20884: gap of unknown length
* 20885 29093: contig of 8209 bp in length
* 29094 29194: gap of unknown length
* 29194 40511: contig of 11316 bp in length
* 40512 40611: gap of unknown length
* 40612 52409: contig of 11798 bp in length
* 52410 52509: gap of unknown length
* 52510 64775: contig of 12266 bp in length
* 64776 64875: gap of unknown length
* 64876 83398: contig of 18523 bp in length
* 83399 83498: gap of unknown length
* 83499 100056: contig of 16558 bp in length
* 100057 100157: gap of unknown length
* 100157 119406: contig of 19250 bp in length.

FEATURES
source
1..119406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-415P10"
/clone_1fb="RPC1 human BAC library 11"
1..4564
/note="assembly_name:Contig3"
misc_feature
4665..12930
/note="assembly_name:Contig50"
misc_feature
13031..20784
/note="assembly_name:Contig31"
misc_feature
20885..29093
/note="assembly_name:Contig32"
misc_feature
29194..40511
/note="assembly_name:Contig33"
misc_feature
40612..52409

/note="assembly_name:Contig34"
misc_feature
52510..64775
/note="assembly_name:Contig35"
misc_feature
64876..83398
/note="assembly_name:Contig36"
misc_feature
83499..100056
/note="assembly_name:Contig37"
misc_feature
100157..119406
/note="assembly_name:Contig38"
BASE COUNT 35588 a 23913 c 23787 g 35137 t 981 others
ORIGIN
Query Match 71.6%; Score 358.2; DB 2; Length 119406;
Best Local Similarity 88.1%; Pred.No.1.6e-107;
Matches 413; Conservative 0; Mismatches 53; Indels 3; Gaps 2;

QY 34 AAAAGCGCTGAGATPACAGCAATTATCTTCACCTGAGACAGAGACTAGCTGATTCC 93
DB 37607 AAAATGATAGAGCTACAGTATGAGACATACAGATGAGATGAGACTAGCTGATTCC 37548
QY 94 TAGCCGACTAAGATCCCTAAGCCTAGCTGGAAGGTGACACGTCACCTTTAAACAC 153
DB 37547 TAGCCGACTAAGATCCCTAAGCCTAGCTGGAAGGTGACACGTCACCTTTAAACAT 37488
QY 154 GGGGCTTGCACTTACTGACACCTGACCAATACAGAGCTCACTAAATGCTAATTAGG 213
DB 37487 GGGGCTTGCACTTACTGACACCTGACCAATACAGAGCTCACTAAATGCTAATTAGG 37428
QY 214 CAATGACAGAGATGAAGAAATGCAATCATCTATTCCTGAGAGACAGAGAGGA 273
DB 37427 CAATGACAGAGATGAAGAAATGCAATCATCTATTCCTGAGAGACAGAGAGGA 37368
QY 274 CAATGACAGAGATGAAGAAATGCAATCATCTATTCCTGAGAGACAGAGAGGA 37333
DB 37367 CAATGACAGAGATGAAGAAATGCAATCATCTATTCCTGAGAGACAGAGAGGA 37309
QY 334 CTTCCTTTGATGAGG--CTGTTTCATGCTATTTCACTATTAATTTGCAACTG 391
DB 37308 CTTCCTTTGATGAGG--CTGTTTCATGCTATTTCACTATTAATTTGCAACTG 37249
QY 392 CACCTTCCTGCTCATCTTTCTTAAGCTGAGCTGAGCTTTGCTCACCCTGCACT 451
DB 37248 CACCTTCCTGCTCATCTTTCTTAAGCTGAGCTGAGCTTTGCTCACCCTGCACT 37189
QY 452 GCTTTTGCACACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500
DB 37188 GCTTTTGCACACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 37140

RESULT 11
AL353584 0/c
WPCOMMENT
Sequence split into 6 fragments LOCUS AL353584 Accession AL353584
Fragment Name Begin End
AL353584_0 1 110000
AL353584_1 100001 210000
AL353584_2 200001 310000
AL353584_3 300001 410000
AL353584_4 400001 510000
AL353584_5 500001 536214
LOCUS AL353584 536214 bp DNA linear HTG 19-JUL-2001
DEFINITION Homo sapiens chromosome X clone RP13-185F3, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL353584
VERSION AL353584.11 GI:15020381
KEYWORDS HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 536214)
AUTHORS Wilson,S.
TITLE Direct Submission

JOURNAL

Submitted (18-JUN-2001) Sanger Centre, Hinxton, Cambridgehire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jul 25, 2001 this sequence version replaced gi:13121369.

COMMENT

Center: Sanger Centre
 Genome Center

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

Project Information

Center project name: bB185F3

Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; 108752; 100% of reads

Chemistry: Dye-terminator ET-amersham; 28% of reads Chemistry:

Dye-terminator Big Dye; 71% of reads

Consensus quality: 167016 bases at least Q40

Consensus quality: 167873 bases at least Q30

Consensus quality: 168523 bases at least Q20

Insert size: 536214; sum-of-contigs

Insert size: 160797; 5.5% error; agarose-fp

Quality coverage: 2.13x in Q20 bases; sum-of-contigs Quality

Coverage: 7.40x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence.

This sequence will be replaced

by the finished sequence as soon as it is available and

the accession number will be preserved.

Location/Qualifiers

1..536214

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/clone="RP13-185F3"

/clone_lib="RPCT-13.1"

/note="assembly_fragment:03802"

misc_feature

1..536214

/note="assembly_fragment:03802"

clone_end:SP6

vector_side:left

clone_end:T7

vector_side:left

BASE COUNT 169781 a 102600 c 103981 g 159850 t 2 others

ORIGIN

Query Match 71.6%; Score 357.8; DB 2; Length 110000;

Best Local Similarity 92.0%; Pred.No.2.2e-107;

Matches 400; Conservative 0; Mismatches 32; Indels 3; Gaps 2;

QY 68 TGAGAGACAGACTAGCTGATTTCTAGGCGAGCTAAGATCCCTAAGCTAGCTGGG 127

DB 38246 TGAGAGACAGACTAGCTGATTTCTAGGCGAGCTAAGATCCCTAAGCTAGCTGGG 38187

QY 128 AGGTGACACAGTCACCTTTAAACAGGGGCTTGCAACTTAGCTCACACCTGACCAATCA 187

DB 38186 AGGTGACAGTCACCTTTAAACAGGGGCTTGCAACTTAGCTCACACCTGACCAATCA 38127

QY 188 GAGAGCTCACTAAATGCTAATAGCAAGAGAGAGATTAAGAAATAGCCATCATCT 247

DB 38126 GAGAGCTCACTAAATGCTAATAGCAAGAGAGAGATTAAGAAATAGCCATCATCT 38067

QY 248 ATTGCTGAGAGACAGAGAGAGCAACAATCGGATATTAACCGAGCATTCGAGCT 307

DB 38066 ATTGCTGAGAGACAGAGAGAGCAACAATCGGATATTAACCGAGCATTCGAGCT 38007

QY 308 GGGCAAGACAGCCCCCTTTGGTCCCTTCCTTTGTATGGAG--CTGTTTCATGCTA 365

DB 38006 GGGCAAGACAGCCCCCTTTGGTCCCTTCCTTTGTATGGAG--CTGTTTCATGCTA 37948

QY 366 TTTCACCTATTAATCTTCAACTGACCTCTGCTCATATTTCTTAAGGCTGAGC 425

DB 37947 TTTCACCTATTAATCTTCAACTGACCTCTGCTCATATTTCTTAAGGCTGAGC 37888

QY 426 TGAGCTTTGCTCAGCGTCCACCACTGCTGTTTGCAACACCGAGACCTGCGCTGACT 485

Db 37887 TGAGCTTTGCTCAGCGTCCACCACTGCTGTTTGCAACACCGAGACCTGCGCTGACT 37828

QY 486 CCCATCCCTTGAT 500

Db 37827 TCCATCCTTCAGAT 37813

RESULT 12

CNS06C7R/c

LOCUS

DEFINITION

Human chromosome 14 DNA sequence BAC R-254B15 of library RPCT-11

from chromosome 14 of Homo sapiens (Human), complete sequence.

ACCESSION

AL390800

VERSION

AL390800.4

KEYWORDS

HTG.

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE


```

/note="HERV16 repeat: matches 4. .664 of consensus"
repeat_region 63927. .64774
/note="HERV16 repeat: matches 716. .1577 of consensus"
repeat_region 64789. .64832
/note="22 copies 2 mer aa 81% conserved"
repeat_region 64916. .65612
/note="HERV16 repeat: matches 1817. .2518 of consensus"
repeat_region 65613. .66886
/note="THE1-INTERNAL repeat: matches 124. .1401 of consensus"
misc_feature 66475. .66484
/note="1372 bp of 18186 transposon (X03123) removed here. This sequence represents the duplicated flanking sequence of the 18186."
70601. .70772
/note="match: GSS: Em:AQ077330"
70734. .71229
/note="match: GSS: Em:AQ628837"
80000. .80512
/note="match: GSS: Em:AQ284023
match: STS: Em:G61704"
complement(81135. .81789)
/note="match: GSS: Em:AQ240981"
81193. .81655
/note="match: GSS: Em:AQ704174"
83625. .89762
/note="L1PA5 repeat: matches 11. .6145 of consensus"
91802. .91895
/note="L1PA3 repeat: matches 6053. .6146 of consensus"
94922. .101052
/note="L1PA7 repeat: matches 11. .6145 of consensus"
102779. .102828
/note="25 copies 2 mer ac 82% conserved"
complement(103577. .104157)
/note="match: GSS: Em:AQ775154"
105177. .105697
/note="match: GSS: Em:B63384"
105842. .106341
/note="match: GSS: Em:AQ522879"
107797. .108270
/note="MER39 repeat: matches 13. .521 of consensus"
116826. .116918
/note="MER91A repeat: matches 1. .94 of consensus"
complement(122839. .123304)
/note="match: GSS: Em:AQ002183"
126653. .126696
/note="22 copies 2 mer ac 93% conserved"
127908. .127950
/note="L1PA5 repeat: matches 6103. .6145 of consensus"
complement(129916. .130692)
/note="match: GSS: Em:A2288149"
complement(130168. .130745)
/note="match: GSS: Em:A2293756"
complement(131672. .132168)
/note="match: GSS: Em:AQ466443"
131821. .131846
/note="13 copies 2 mer ta 96% conserved"
132163. .132600
/note="match: GSS: Em:AQ829322"
133493. .133664
/note="L1R33 repeat: matches 30. .202 of consensus"
136095. .136245
/note="L1MA7 repeat: matches 6142. .6289 of consensus"
complement(140979. .141356)
/note="match: STS: Em:HSC28H08"
146003. .146304
/note="L1R33 repeat: matches 1. .310 of consensus"
146351. .146421
/note="L1R33 repeat: matches 448. .516 of consensus"
157021. .157123
/note="MER34 repeat: matches 76. .173 of consensus"
159023. .159689
/note="HERV16 repeat: matches 562. .1217 of consensus"

```

```

repeat_region 159978. .160033
/note="HERV16 repeat: matches 1217. .1272 of consensus"
repeat_region 164301. .166134
/note="L1PA14 repeat: matches 1841. .3688 of consensus"
repeat_region 166458. .166509
/note="26 copies 2 mer ga 73% conserved"
repeat_region 166568. .169037
/note="L1PA14 repeat: matches 3681. .6148 of consensus"
misc_feature 169389. .169794
/note="match: GSS: Em:AQ092454"
misc_feature 170746. .170746
complement(170139. .170746)
/note="match: GSS: Em:AQ389950"
misc_feature 170286. .170416
complement(170286. .170416)
/note="match: GSS: Em:AQ835692"
BASE COUNT 53940 a 32801 c 34298 g 49707 t
ORIGIN
Query Match 71.6%; Score 357.8; DB 9; Length 170746;
Best Local Similarity 92.0%; Pred. No. 2.3e-107;
Matches 400; Conservative 0; Mismatches 32; Indels 3; Gaps 2;
QY 68 TGAGAGCAGGACTGAGTGGATTTCCTAGGCGCACTAAGATCCCTAAGCTTAGCTGGGA 127
Db 38246 TGAGAGCAGGACTGAGTGGATTTCCTAGGCGCACTAAGATCCCGAAGCTTAGCTGGGA 38187
QY 128 AGGTGACACAGTCCACCTTTAAACAGGGGCTTGCAACTTAGCTCACACCTGACCAATCA 187
Db 38186 AGGTGACATCCATCCATTTTAAACAGGGGCTTGCAACTTAGCTCACACCTGACCAATCA 38127
QY 188 GAGAGCTCACTAAATGCTATTATAGGCAAGAGAGTAAAGAAATAGCAATCATCT 247
Db 38126 GAGAGCTCACTAAATGCTATTATAGGCAAGAGAGTAAAGAAATAGCAATCATCT 38067
QY 248 ATTGCTGAGAGCAGCAGAGAGGAGACAAATCGGATTTAAACCGAGCATTCGAGCT 307
Db 38066 ATTGCTGAGAGCAGCAGAGAGGAGAAAGATCGGATTTAAACCGAGCATTCGAGCC 38007
QY 308 GGCAGCAGAGCCCGCTTTGGGTCCTTCCTTTGATGGAG--CTGTTTCATGCTA 365
Db 38006 GGCAGCAGCA-ACCCCTTTGGGTCCTTCCTTTGATGGAGCTGTTTCATGCT 37948
QY 366 TTTCACCTATTAAATCTTGCAACTGCACTCTTGCTGCATGTTCTTAAGGCTCGAGC 425
Db 37947 TTTCACCTATTAAATCTTGCAACTGCACTCTTGCTGTTGTTTACGGCTGAGC 37888
QY 426 TGAGCTTTGCTCACCGTCCACCACTGCTGTTGGCAACCCGAGACCTCGCTGACT 465
Db 37887 TGAGCTTTGCTCGCACTGCTGCTGTTGGCGCGCTCGAGACCCGCGCTGACT 37828
QY 486 CCCATCCCTCGGAT 500
Db 37827 TCCATCCCTCGGAT 37813

```

RESULT 14
AC093531/c 163803 bp DNA linear PRI 16-NOV-2001
LOCUS Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
DEFINITION AC093531
AC093531.2 GI:16945981
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 163803)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 163803)
DOE Joint Genome Institute.
AUTHORS
TITLE
JOURNAL
Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint

REFERENCE 3 (bases 1 to 163803)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Nov 16, 2001 this sequence version replaced gi:15383820.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing completed at Stanford Human Genome Center
 www.hngc.stanford.edu
 Quality: Phrap Quality >=40 99.8% cf Sequence;
 Estimated Total Number of Errors is 0.3.
 Location/Qualifiers
 1. 163803
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-40517"

BASE COUNT 54797 a 31372 c 29086 g 48548 t
 ORIGIN

Query Match 71.5%; Score 357.4; DB 9; Length 163803;
 Best Local Similarity 91.2%; Pred. No. 3.1e-107;
 Matches 402; Conservative 0; Mismatches 36; Indels 3; Gaps 2;

QY 62 TGCACCTAGAGACGACGCTGATTTCTAGGCTGACCTAAGAACTCCCTAGCCTAG 121
 DB 120271 TACAAATTAGACACGACGCTGATTTCTAGGCTGACCTAAGAACTCCCTAGCCTAG 120212
 QY 122 CTGGGAGGTGACGACGCTGACCTTTAAACAGGGGC.TGCAACTTAGCTCAACCTGAC 181
 DB 120211 CTGGGAGGTGACGCTGACCTTTAAACAGGGGC.TGCAACTTAGCTCAACCTGAC 120152
 QY 182 CAATCAGAGAGCTCACTAATGCTAATTAGGCAAGACAGAGGTAAGAAATGCCCCA 241
 DB 120151 CAATCAGAGAGCTCACTAATGCTAATCAGCAAAACAGAGGCAAGAAATGCCCCA 120092
 QY 242 TCATCTATTGCTGAGACGACGAGGAGCAACATCGGATATATAACCCAGGCAAT 301
 DB 120091 TCATCTATTGCTGAGACGAGGAGGAGCAACATCGGATATATAACCCAGGCAAT 120032
 QY 302 CGAGCTGGACACGACGCCCCCTTTGGGTCCCTTCTTTGTATGGAG--CTGTTTTC 359
 DB 120031 TGAGCTGGAAATGGCA-ACCCCTTTGGGTCCCTTCTTTGTATGGAGCTGTTTTC 119973
 QY 360 ATGCTATTCTCTATTAAATCTTGCAATGCACTCTCTGCTGCTATGTTCTTAAGGC 419
 DB 119972 ACTCTATTCTCTATTAAATCTTGCAATGCACTCTCTGCTGCTATGTTCTTAAGGC 119913
 QY 420 TCGAGCTGAGCTTTTGTACAGGCTCCAGCAGCTGTTTGCACACGCGACAGCTGCGC 479
 DB 119912 TCGAGCTGAGCTTTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 119853
 QY 480 CTGACTCCCATCCCTCTGGAT 500
 DB 119852 CTGACTTCATCCCTCCAGAT 119832

RESULT 15
 AC010888 191863 bp DNA linear PRI 07-NOV-2001
 LOCUS Homo sapiens BAC clone RP11-418N16 from 2, complete sequence.
 ACCESSION AC010888
 VERSION AC010888.12 GI:13992760
 KEYWORDS HTG.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 191863)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 191863)
 AUTHORS Margulis, C., Abbott, A. and Sawyer, C.
 TITLE The sequence of Homo sapiens BAC clone RP11-418N16
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 191863)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (25-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 191863)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAY-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 5 (bases 1 to 191863)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 6 (bases 1 to 191863)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On May 8, 2001 this sequence version replaced gi:13431186.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0418N16

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 All regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 The RP11-418N16 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frangen, E., Tateo, M., Catanesi, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)

VECTOR: pBAC3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-434M17, 200 bp overlap. Actual start of this clone is at base position 195 of RP11-418N16; actual end is at base position 191863 of RP11-418N16.

Db 171735 TCGAGCTGAGCTTTGCTCGCCATCCACCAAGACTGTTGGCGCCCTCGACACCCGCTG 171794

Qy 480 CTGACTCCCATCCCTCTGGAT 500

Db 171795 CTGACTCCCATCCCTCTCGAGAT 171815

Search completed: April 17, 2003, 06:37:59
Job time : 2080.07 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 02:30:00 ; Search time 180.426 Seconds
(without alignments)
6240.778 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Perfect score: 500

Sequence: 1 cccctggcgcggtccctcctt.....tgatcccatccctctgat 500

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_101002.*

```
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	10499	21	ABN97929
2	500	100.0	56093	24	ABL61744
3	435	87.0	711	21	ABN97946
4	413.4	82.7	7466	23	AA568626
5	410	82.0	783	20	AA525669
6	383.8	76.8	711	21	ABN97947
7	348.2	69.6	1329	19	AAV43219
8	348.2	69.6	1329	20	AA529704
9	326.8	65.4	1393	22	AA531002

10	306.8	61.4	893	23	AA565964
11	296.2	59.2	849	22	AA531000
12	295.6	59.1	1478	22	AA587568
13	292.4	58.5	808	23	AA588392
14	290.8	58.2	1243	23	AA584189
15	287.6	57.5	2629	22	AAK87491
16	286.4	57.3	2781	22	AA555630
17	284.8	57.0	2346	21	AA577526
18	284.8	57.0	2462	21	AA559468
19	284.6	56.9	2300	23	AA592680
20	283.8	56.8	2930	24	AA524195
21	283.2	56.6	3372	21	AA525663
22	283.2	56.6	3372	21	AA559213
23	283.2	56.6	7582	20	AA525665
24	283.2	56.6	7582	21	AA559215
25	282.2	56.4	3831	23	AA571727
26	282.2	56.4	5154	23	AA567609
27	282.2	56.4	8279	23	AA576474
28	282.2	56.4	8294	23	AA584209
29	281.6	56.3	2782	20	AA525661
30	281.6	56.3	2782	21	AA559211
31	281.6	56.3	2782	22	AA520069
32	280.8	56.2	1165	23	AA572721
33	280.6	56.1	1136	20	AA525660
34	280.6	56.1	1136	21	AA559210
35	280	56.0	2942	22	AA577313
36	279.2	55.8	17758	23	AA584138
37	277.6	55.5	416	23	AA565963
38	270.4	54.1	635	19	AAV43215
39	270.4	54.1	635	20	AA529702
40	270.4	54.1	2030	21	AA563826
41	261.4	52.3	410	24	ABN94040
42	261.4	52.3	410	24	ABK64210
43	259.6	51.9	448	23	AA587567
44	259.6	51.9	583	23	AA588391
45	259.2	51.8	541	22	AA562368

ALIGNMENTS

```
RESULT 1
ID ABN97929 standard; DNA; 10499 BP.
XX
XX ABN97929;
AC
XX 01-AUG-2002 (first entry)
DT
XX
XX Human retroviral sequence HERV-7q.
DE
XX
XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
XX
XX Human retrovirus.
OS
XX
XX WO967395-A1.
PN
XX
XX 29-DEC-1999.
PD
XX
XX 23-JUN-1999; 99WO-FR01513.
PF
XX
XX 23-JUN-1998; 98FR-0007920.
PR
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA
XX
XX Alliel PM, Perin J, Rieger F;
PI
XX
XX WPI; 2000-160587/14.
DR
XX
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -
```

XX Claim 3; Fig 1; 225bp; French.
XX
CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;

Query Match 100.0%; Score 500; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 5; le-164;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGGGGGCGGGCTTCTTCTGGGATGAGGGCAAAAGCTGAGATACAGCAATTATC 60
DB 1 CCTGGGGGCGGGCTTCTTCTGGGATGAGGGCAAAAGCTGAGATACAGCAATTATC 60

QY 61 TTGGAACGTGAGACAGGACTGAGTGTGATTTCTTAGGCGAGCTAAGATCCCTAAGCCTA 120
DB 61 TTGGAACGTGAGACAGGACTGAGTGTGATTTCTTAGGCGAGCTAAGATCCCTAAGCCTA 120

QY 121 GCTGGGAAGGTGACCACTCCACTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGA 180
DB 121 GCTGGGAAGGTGACCACTCCACTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGA 180

QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCA 240
DB 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCA 240

QY 241 ATCATCTATTGCTGAGAGCAGCAGAGAGGGAACAACATGGGGATATTAACCCAGGCAT 300
DB 241 ATCATCTATTGCTGAGAGCAGCAGAGAGGGAACAACATGGGGATATTAACCCAGGCAT 300

QY 301 TCGAGCTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTGTATGGAGCTGTTTCA 360
DB 301 TCGAGCTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTGTATGGAGCTGTTTCA 360

QY 361 TGTATTTCACCTATTAAATCTTGCAACTGCACTCTTGTGTCATGTTCTTAACGGCT 420
DB 361 TGTATTTCACCTATTAAATCTTGCAACTGCACTCTTGTGTCATGTTCTTAACGGCT 420

QY 421 CGAGCTGAGCTTTGGCTCAGCGTCCAGACAGCTGTTTCCACACCGGCAACCGCGCGC 480
DB 421 CGAGCTGAGCTTTGGCTCAGCGTCCAGACAGCTGTTTCCACACCGGCAACCGCGCGC 480

QY 481 TGACTCCCATCCCTCTGGAT 500
DB 481 TGACTCCCATCCCTCTGGAT 500

RESULT 2
ABL61744
ID ABL61744 standard; DNA; 56093 BP.
XX
AC ABL61744;
XX
DT 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:81.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM Stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
CY cytosstatic; gene therapy; anti-neoplastic; Wilms' tumor; adenocarcinoma;

KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 22-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.

(AVAL-) AVALON PHARM.
XX
PA
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
DR
XX
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 81; 44bp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL770110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytosstatic

CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophagial, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.

XX Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 other;

XX Query Match 100.0%; Score 500; DB 24; Length 56093;

XX Best Local Similarity 100.0%; Pred. No. 1.3e-163;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTGGGGCGGGCTTCTTCTTGGGATGAGGCAAAACGCTGAGATACGCAATTATC 60
DB 28001 CCTTGGGGCGGGCTTCTTCTTGGGATGAGGCAAAACGCTGAGATACGCAATTATC 28060
QY 61 TTGCACTGAGAGACAGGACTAGGCTGATTTCTTGAAGGCGCACTAAGATCCCTAAGCCTA 120
DB 28061 TTGCACTGAGAGACAGGACTAGGCTGATTTCTTGAAGGCGCACTAAGATCCCTAAGCCTA 28120
QY 121 GCTGGGAAGGTGACCAAGTCCACCTTTTAAACAGCGGGCTTGCACTTACCTACACCTGA 180
DB 28121 GCTGGGAAGGTGACCAAGTCCACCTTTTAAACAGCGGGCTTGCACTTACCTACACCTGA 28180
QY 181 CCAATCAGAGAGCTGACTAAATGCTAATTAGGCAAGAGAGATTAAGAAATAGCCA 240
DB 28181 CCAATCAGAGAGCTGACTAAATGCTAATTAGGCAAGAGAGATTAAGAAATAGCCA 28240
QY 241 ATCATCTATTGCTGAT 300
DB 28241 ATCATCTATTGCTGAT 28300
QY 301 TCGAGCTGGCAACAGAT 360
DB 28301 TCGAGCTGGCAACAGAT 28360
QY 361 TGTATTTTCACTATTAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 420
DB 28361 TGTATTTTCACTATTAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 28420
QY 421 CGAGCTGAGCTTTTGTCTACCGTCCACAGCTGTTTGGCCACACCGGCAACCTGCGGC 480
DB 28421 CGAGCTGAGCTTTTGTCTACCGTCCACAGCTGTTTGGCCACACCGGCAACCTGCGGC 28480
QY 481 TGACTCCCATCCCTCTGGAT 500
DB 28481 TGACTCCCATCCCTCTGGAT 28500

RESULT 3

ABN97946

ID ABN97946 standard; DNA; 711 BP.

XX AC ABN97946;

XX DT 01-AUG-2002 (first entry)

XX DE Human retroviral sequence R1.

XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

XX KM multiple sclerosis; ds.

XX OS Human retrovirus.

XX PN WO967395-A1.

XX PD 29-DEC-1999.

PF 23-JUN-1999; 99WO-FR01513.

XX 23-JUN-1998; 98FR-0007920.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX PI Alliel PM, Perin J, Rieger F;

XX WPI; 2000-160587/14.

XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,

PT used for diagnosis, treatment and prevention of autoimmune and

PS Claim 3; Fig 3; 225pp; French.

CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.

XX Sequence 711 BP; 170 A; 204 C; 162 G; 175 T; 0 other;

XX Query Match 87.0%; Score 435; DB 21; Length 711;

XX Best Local Similarity 100.0%; Pred. No. 7.1e-142;

Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ACTGAGAGACAGAGCTAGCTGATTTCTTGAAGGCGCACTAAGATCCCTAAGCTAGCTG 125
DB 1 ACTGAGAGACAGAGCTAGCTGATTTCTTGAAGGCGCACTAAGATCCCTAAGCTAGCTG 60
QY 126 GAAGGTGACCAAGTCCACCTTTTAAACAGGGGCTTGCACTTACCTACACCTGACCAAT 185
DB 61 GAAGGTGACCAAGTCCACCTTTTAAACAGGGGCTTGCACTTACCTACACCTGACCAAT 120
QY 186 CAGAGAGCTACTAAATGCTAATTAGGCAAGAGAGAGAGAGAGAGAGAGATAGCAATCAT 245
DB 121 CAGAGAGCTACTAAATGCTAATTAGGCAAGAGAGAGAGAGAGAGAGATAGCAATCAT 180
QY 246 CTATTGCTGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 305
DB 181 CTATTGCTGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 240
QY 306 CTGGCAACAGAT 365
DB 241 CTGGCAACAGAT 300
QY 366 TTTCACTTATTAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 425
DB 301 TTTCACTTATTAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 360
QY 426 TGAGCTTTTGTCTACCGTCCACAGCTGTTTGGCCACACCGGCAACCTGCGGC 485
DB 361 TGAGCTTTTGTCTACCGTCCACAGCTGTTTGGCCACACCGGCAACCTGCGGC 420
QY 486 CCATCCCTCTGGAT 500
DB 421 CCATCCCTCTGGAT 435

RESULT 4

AA568626

ID AA568626 standard; cDNA; 7466 BP.

XX

AC AAS68626;
 XX 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic proceain #4430.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX MO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX P-PSDB; ABG04439.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 1; SEQ ID No 4430; 103bp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX SQ Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;
 XX
 XX Query Match 82.7%; Score 413.4; DB 23; Length 7466;
 XX Best Local Similarity 98.2%; Pred. No. 1e-133;
 XX Matches 429; Conservative 0; Mismatches 6; Indels 2; Gaps 1;
 XX
 QY 66 ACTGAGAGACAGACTAGTGTGATTCCTAGGCGGACTAGATCCCTAGGCTAGCTGG 125
 DB 3716 ACTGAGAGACAGACTAGTGTGATTCCTAGGCGGACTAGATCCCTAGGCTAGCTGG 3775
 QY 126 GAAGTGACACGCTCCACCTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAAT 185
 DB 3776 GAAGTGACACGCTCCACCTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAAT 3835
 QY 186 CAGAGAGCTCACTAAATGCTATTAGGCAAGACAGGCGGTAAAGAAATAGCAATCAT 245
 DB 3836 CAGAGAGCTCACTAAATGCTATTAGGCAAGACAGGCGGTAAAGAAATAGCAATCAT 3895

QY 246 CTATTGCTGAGACAGACAGAGGGAACAATCGGATATTAACCCAGGCAATTCGAG 305
 DB 3896 TTATTGCTGAGACAGACAGAGGGAACAATCGGATATTAACCCAGGCAATTCGAG 3955
 QY 306 CTGGCAACAGACAGCCCGCTTTGGGTCCCTTCCTTTGTATAGGAG--CTGTTTCAATGC 363
 DB 3956 CTGGCAACAGACAGCCCGCTTTGGGTCCCTTCCTTTGTATAGGAGCTGTTTCAATGC 4015
 QY 364 TATTCACTATTAAATCTTGCAACTGACTCTTCTGTCATGTTCTTACGAGCTGA 423
 DB 4016 TATTCACTATTAAATCTTGCAACTGACTCTTCTGTCATGTTCTTACGAGCTGA 4075
 QY 424 GCTGAGCTTTTGTCTCACCCTGACCACTGCTTTTCCACACCGAGCACTGCCGCTGA 483
 DB 4076 GCTGAGCTTTTGTCTCACCCTGACCACTGCTTTTCCACACCGAGCACTGCCGCTGA 4135
 QY 484 CTCCCATCCCTCTGGAT 500
 DB 4136 CTCCCATCCCTCTGGAT 4152

RESULT 5
 AAX25669
 ID AAX25669 standard; cDNA to mRNA; 783 BP.
 XX
 XX AAX25669;
 XX
 XX 21-MAY-1999 (first entry)
 XX
 XX Human endogenous retrovirus W long terminal repeat region.
 DE
 XX Clone; human endogenous retrovirus; genome; autoimmune disease;
 XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 XX disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
 OS Human endogenous retrovirus.
 XX
 XX WO9902696-A1.
 XX 21-JAN-1999.
 XX
 XX 06-JUL-1998; 98WO-FR01442.
 XX
 XX 07-JUL-1997; 97FR-0008815.
 XX
 XX (INMR) BIO MERIEUX.
 XX
 XX Beseme F, Blond JL, Bouton O, Maillet F, Mandrand B;
 XX WPI; 1999-120897/10.
 XX
 XX New nucleic acid sequences from human endogenous retrovirus-W -
 PT expressed exclusively in placenta and useful in diagnosis and
 PT therapy of autoimmune disease, and abnormal or failed pregnancy
 PT
 XX
 XX Claim 1; Page 83; 106bp; French.
 XX
 XX This sequence represents the long terminal repeat (LTR) region of the
 CC human endogenous retrovirus (HERV) W genome. The nucleic acids, their
 CC fragments or peptides encoded by them are markers of autoimmune disease
 CC (e.g. multiple sclerosis, rheumatoid polyarthritis, disseminated lupus
 CC erythematosus, insulin-dependent diabetes and related pathologies) and
 CC of abnormal or unsuccessful pregnancy and can be used as chromosomal
 CC markers for susceptibility to these conditions, or proximity markers
 CC of genes associated with this susceptibility.
 XX
 XX SQ Sequence 783 BP; 173 A; 213 C; 166 G; 180 T; 51 other;
 XX
 XX Query Match 82.0%; Score 410; DB 20; Length 783;
 XX Best Local Similarity 93.8%; Pred. No. 4.5e-133;
 XX Matches 408; Conservative 24; Mismatches 1; Indels 2; Gaps 1;
 XX

QY 68 TGAAGACAGAGCTAGCTGATTTCTGAGGCCGACTAAGAAATCCCTAAGCTAGCTGGGA 127
 DB 1 TGACAGACAGAGCTAGCTGATTTCTGAGGCCGACTAAGAAATCCCTAAGCTAGCTGGGA 60
 QY 128 AGGAGACAGAGCTAGCTGATTTCTGAGGCCGACTAAGAAATCCCTAAGCTAGCTGGGA 187
 DB 61 AGGAGACAGAGCTAGCTGATTTCTGAGGCCGACTAAGAAATCCCTAAGCTAGCTGGGA 120
 QY 188 GAGAGCTAGCTAGCTGATTTCTGAGGCCGACTAAGAAATCCCTAAGCTAGCTGGGA 247
 DB 121 GAGAGCTAGCTAGCTGATTTCTGAGGCCGACTAAGAAATCCCTAAGCTAGCTGGGA 180
 QY 248 ATTGCTAGAGAGCAGAGAGGAGGAGCAATCGGATATTAACCCAGGATTCGAGCT 307
 DB 181 ATTGCTAGAGAGCAGAGAGGAGGAGCAATCGGATATTAACCCAGGATTCGAGCT 240
 QY 308 GGAACAGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 365
 DB 241 GGAACAGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 300
 QY 366 TTTGACTATTAATCTTGAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 425
 DB 301 TTTGACTATTAATCTTGAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
 QY 426 TGAAGCTTTGCTCAACGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 485
 DB 361 TGAAGCTTTGCTCAACGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
 QY 486 CCCATCCCTCTGGAT 500
 DB 421 CCCATCCCTCTGGAT 435

RESULT 6

ID AEN97947 standard; DNA; 711 BP.
 AC AEN97947;
 XX
 DT 01-AUG-2002 (first entry)
 DE Human retroviral sequence RIF.
 XX
 KM Autoimmune disease; HERV-7g; chromosome 7g; immunotherapy;
 XX multiple sclerosis; ds.
 OS Human retrovirus.
 XX
 PN WO967395-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 23-JUN-1999; 99WO-FR01513.
 XX
 PR 23-JUN-1998; 98FR-0007920.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Alliel PM, Perin J, Rieger F;
 XX
 DR WPI; 2000-160587/14.
 XX
 PT New nucleic acid sequences of human endogenous retrovirus, HERV-7g,
 PT used for diagnosis, treatment and prevention of autoimmune and
 PT neurological diseases -
 XX
 PS Claim 3; Fig 3; 225BP; French.
 XX
 CC The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7g, which is located on chromosome 7g.
 CC Regulatory elements associated with HERV-7g may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/

CC curative). HERV-7g peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7g peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention.
 XX
 SQ Sequence 711 BP; 171 A; 209 C; 160 G; 171 T; 0 other;

Query Match 76.8%; Score 383.8; DB 21; Length 711;
 Best Local Similarity 95.4%; Pred. No. 6.6e-124;
 Matches 417; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

QY 66 ACTGAG 125
 DB 1 ACTGAG 60
 QY 126 GAAAGTACACAG 185
 DB 61 GAAAGTACACAG 120
 QY 186 CAGAGAGCTCACTAAATGCTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245
 DB 121 CAGAGAGCTCACTAAATGCTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 QY 246 CTATTTGCTGAG 305
 DB 181 CTATTTGCTGAG 240
 QY 306 CTGCAACAG 363
 DB 241 CTGCAACAG 299
 QY 364 TATTTCACTATTAATCTTGAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 423
 DB 300 TATTTCACTATTAATCTTGAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 359
 QY 424 GCTGAGCTTTGCTCAACGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 483
 DB 360 GCTGAGCTTTGCTCAACGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 419
 QY 484 CTCCATCCCTCTGGAT 500
 DB 420 CTCCATCCCTCTGGAT 436

RESULT 7

ID AAV43219 standard; cDNA; 1329 BP.
 AC AAV43219;
 XX
 DT 29-DEC-1998 (first entry)
 DE Multiple sclerosis associated retrovirus fragment 6.
 XX
 KM Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
 KM gag gene; env gene; rheumatoid arthritis-associated virus; ss.
 XX
 OS Multiple sclerosis associated retrovirus.
 XX
 FH Key Location/Qualifiers
 FT CDS 2..490
 FT /*tag= a
 FT /product= "Encodes protein AAV71069"
 FT /transl_except= (pos:77-79, appears to code for a
 FT stop codon)
 FT /transl_except= (pos:125-127, appears to code for a
 FT stop codon)
 FT /transl_except= (pos:137-139, appears to code for a
 FT stop codon)

XX MO9823755-A1.
XX 04-JUN-1998.
XX 26-NOV-1997; 97WO-1B01482.
XX 26-NOV-1996; 96US-0756429.
XX (INMR) BIO MERIEUX.
XX Bedin F, Beeseme F, Jolivet-Reynaud C, Komurian-Pradel F,
XX Mandrand B, Paranhos-Baccala G, Perron H;
XX WPI; 1998-322732/28.
XX P-PSDB; AAW71069.
XX
XX New nucleic acid from retroviruses - useful for diagnosis,
XX prevention and treatment of, e.g. multiple sclerosis
XX
XX Disclosure; Pages 187-188; 286pp; English.
XX
XX The present sequence represents a multiple sclerosis (MS) associated
XX retrovirus (MSRV) genomic fragment used in the method of the
XX invention. The invention provides complete or partial genomic
XX sequences of the MSRV-1 pol gene, gag gene, and env gene, and
XX polypeptides encoded by these genes. The invention also provides
XX antibodies raised against the polypeptides. The genomic sequences,
XX polypeptides and antibodies are also claimed useful for diagnosing,
XX infection by MS and rheumatoid arthritis-associated viruses, and also
XX for prevention and treatment of infection with these viruses.
XX
XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 69.6%; Score 348.2; DB 19; Length 1329;
Best Local Similarity 90.2%; Pred. No. 3e-111;
Matches 394; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

QY 66 ACTGAGAGACAGAGCTAGCTGATTTCTTAGGCGGACPAAGATCCCTAAGCTTAGCTG 125
DB 547 ACTGAGAGACAGAGCTAGCTGATTTCTTAGGCGGACPAAGATCCCTAAGCTTAGCTG 606
QY 126 GAAGGTGACACGCTCCACCTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAAT 185
DB 607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCACTTAGCTCACACCGACCAAT 666
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGTAAAGAAATAGCCATCAT 245
DB 667 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGTAAAGAAATAGCCATCAT 726
QY 246 CTATTGCTGAGAGCAGACAGAGAGGAGCAACATCGCATTAATCCAGGCAATTGAG 305
DB 727 CTATTGCTGAGAGCAGACAGAGAGGAGCAACATCGCATTAATCCAGGCAATTGAG 786
QY 306 CTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCTTTGATGGAG--CTGTTTTCAATGC 363
DB 787 CCAGCAACAGCA-ACCCCCCTTTGGGTCCCTTCCTTCATGATGGAGCTGTGTTTCACTGC 845
QY 364 TATTTCACCTATTAAATCTTGCACTGCACTTTCTGTCTCATGTTTCTTACGGCTGCA 423
DB 846 TATTTCACCTATTAAATCTTGCACTGCACTTTCTGTCTCATGTTTCTTACGGCTGCA 905
QY 424 GCTGAGCTTTTGCTCAGCGTCCACCACTGCTGTTTGGCACCAGCGAGACCTGCGCTGA 483
DB 906 GCTGAGCTTTTGCTGCGCATCCACCACTGCTGTTTGGCACCGTACAGACCGGCTGCTGA 965
QY 484 CTCCCATCCCTCTGAT 500
DB 966 CTTCATCCCTTTGAT 982

RESULT 8
AA29704

ID AAX29704 standard; DNA; 1329 BP.
XX AAX29704;
XX 08-JUN-1999 (first entry)
XX Clone 5M6 from MSRV-1.
XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
XX rheumatoid polyarthritis; ss.
XX
XX Multiple sclerosis related virus type 1.
XX
XX FR2765588-A1.
XX 08-JAN-1999.
XX
XX 07-JUL-1997; 97FR-0008816.
XX 07-JUL-1997; 97FR-0008816.
XX 07-JUL-1997; 97FR-0008816.
XX (INMR) BIO MERIEUX.
XX
XX WPI; 1999-098275/09.
XX P-PSDB; AAW99554.
XX
XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
XX multiple sclerosis or rheumatoid polyarthritis
XX
XX Claim 1; Page 39-40; 83pp; French.
XX
XX This sequence represents clone 5M6 from a novel multiple sclerosis
XX related virus type 1 (MSRV1). The sequence can be used in diagnostic,
XX prophylactic or therapeutic compositions to inhibit expression of a
XX multiple sclerosis related virus and/or virus associated with
XX rheumatoid polyarthritis.
XX
XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 69.6%; Score 348.2; DB 20; Length 1329;
Best Local Similarity 90.2%; Pred. No. 3e-111;
Matches 394; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

QY 66 ACTGAGAGACAGAGCTAGCTGATTTCTTAGGCGGACPAAGATCCCTAAGCTTAGCTG 125
DB 547 ACTGAGAGACAGAGCTAGCTGATTTCTTAGGCGGACPAAGATCCCTAAGCTTAGCTG 606
QY 126 GAAGGTGACACGCTCCACCTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAAT 185
DB 607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCACTTAGCTCACACCGACCAAT 666
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAGAATAGCCATCAT 245
DB 667 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAGAATAGCCATCAT 726
QY 246 CTATTGCTGAGAGCAGACAGAGAGGAGCAACATCGGATTAATCCAGGCAATTGAG 305
DB 727 CTATTGCTGAGAGCAGACAGAGAGGAGCAACATCGGATTAATCCAGGCAATTGAG 786
QY 306 CTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCTTTGATGGAG--CTGTTTTCAATGC 363
DB 787 CCAGCAACAGCA-ACCCCCCTTTGGGTCCCTTCCTTCATGATGGAGCTGTGTTTCACTGC 845
QY 364 TATTTCACCTATTAAATCTTGCACTGCACTTTCTGTCTCATGTTTCTTACGGCTGCA 423
DB 846 TATTTCACCTATTAAATCTTGCACTGCACTTTCTGTCTCATGTTTCTTACGGCTGCA 905
QY 424 GCTGAGCTTTTGCTCAGCGTCCACCACTGCTGTTTGGCACCAGCGAGACCTGCGCTGA 483
DB 906 GCTGAGCTTTTGCTGCGCATCCACCACTGCTGTTTGGCACCGTACAGACCGGCTGCTGA 965
QY 484 CTCCCATCCCTCTGAT 500

AA65964;
13-FEB-2002 (first entry)
DNA encoding novel human diagnostic protein #1768.
Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.
Homo sapiens.
W0200175067-A2.
11-OCT-2001.
30-MAR-2001; 2001WO-US08631.
31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
(HYSE-) HYSEQ INC.
Dmanac RT, Liu C, Tang YT;
WPI; 2001-639362/73.
P-PSDB; ABG01777.
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
Claim 1; SEQ ID No 1768; 103pp; English.
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pat_sequences.
Sequence 893 BP; 223 A; 264 C; 202 G; 204 T; 0 other;
Query Match 61.4%; Score 306.8; DB 23; Length 893;
Best Local Similarity 87.2%; Pred. No. 8,1e-97;
Matches 389; Conservative 0; Mismatches 37; Indels 20; Gaps 4;
QY 66 ACTGAGACAGAGCTAGTGGATTTCTTAGGCCGACTAGAGATCCCTAGGCTAGCTGG 125
DB 236 ACTGAGACAGAGCTAGTGGATTTCTTAGGCCGACTAGAGATCCCTAGGCTAGCTGG 295
QY 126 GAAGTGACACAGCTTAAACACCGGGCTTGCA/CTTAGCTCACACCTGACCAAT 185
DB 236 GAAGTGACACAGCTTAAACACCGGGCTTGCA/CTTAGCTCACACCTGACCAAT 355
QY 186 C-----AGAGAGCTACTAAATGCTAATTAGGCTAAGACAGAGAGTAAAGAAAT 236
DB 356 CAGGTAAAGAAAGAGAGCCGCTAAATGCTAATTAGGCTAAGACAGAGTAAAGAAAT 415

QY 237 GCCATCATCTATTGCTGCTGAGACACAGAGGAGAACATCGGATATTAACCCAG 296
DB 416 GCCATCATCTATTGCTGCTGAGACACAGGAGGAGCAATGATCAGATATAACCCAG 475
QY 297 GCATTGAGCTGCAACAGACAGCCCTTTGGGTCCCTTCCCTTTGATAGGAGCTGTT 356
DB 476 GCATTGAGCCGGCAAC-GACTACCTCTTTGGGTCCCTTCCCTTTGATAGGAGCTG 534
QY 357 TTGATGCTATTTCAGCTATTAATCTTGCACTG--CACTCTTGTGTCATGTTTCTT 414
DB 535 T-----TTTCACTATTAATCTTGCACTGCACTCTTGTGTCATGTTTCTT 586
QY 415 ACGGCTGAGCTAGGCTTTTGTCTCACCGTCACACACTGCTTGGCACACCGACAGCC 474
DB 587 ACGGCTGAGCTAGGCTTTTGTCTCACCGTCACACACTGCTTGGCACACCGACAGCC 646
QY 475 TGGCGTGAAGCTCCCATCCCTCTGGAAT 500
DB 647 CATGCTGAAGCTCCCATCCCTCTGGAAT 672
RESULT 11
AAS31000
ID AAS31000 standard; cDNA; 849 BP.
AC AAS31000;
XX 04-DEC-2001 (first entry)
DT Human diagnostic and therapeutic polynucleotide (DIRHP) #15.
XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
respiratory disorder; ss.
XX Homo sapiens.
OS W0200162927-A2.
PN 30-AUG-2001.
PD 21-FEB-2001; 2001WO-US06059.
PF 21-FEB-2001; 2000US-0184693.
XX 24-FEB-2000; 2000US-0184693.
XX 24-FEB-2000; 2000US-0184698.
XX 24-FEB-2000; 2000US-0184768.
XX 24-FEB-2000; 2000US-0184769.
XX 24-FEB-2000; 2000US-0184770.
XX 24-FEB-2000; 2000US-0184771.
XX 24-FEB-2000; 2000US-0184772.
XX 24-FEB-2000; 2000US-0184773.
XX 24-FEB-2000; 2000US-0184774.
XX 24-FEB-2000; 2000US-0184776.
XX 24-FEB-2000; 2000US-0184777.
XX 24-FEB-2000; 2000US-0184777.
XX 24-FEB-2000; 2000US-0184813.
XX 24-FEB-2000; 2000US-0184837.
XX 24-FEB-2000; 2000US-0184841.
XX 24-FEB-2000; 2000US-0185213.
XX 24-FEB-2000; 2000US-0185216.
XX 12-MAY-2000; 2000US-0203785.
XX 15-MAY-2000; 2000US-0204226.
XX 16-MAY-2000; 2000US-0204525.
XX 16-MAY-2000; 2000US-0204821.
XX 16-MAY-2000; 2000US-0204908.
XX 16-MAY-2000; 2000US-0205232.
XX 17-MAY-2000; 2000US-0204815.
XX 17-MAY-2000; 2000US-0204863.
XX 17-MAY-2000; 2000US-0205221.
XX 17-MAY-2000; 2000US-0205285.

CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1478 BP; 389 A; 337 C; 364 G; 380 T; 0 other;
SQ
Query Match 59.1%; Score 295.6; DB 23; Length 1478;
Best Local Similarity 85.7%; Pred. No. 9,4e-93;
Matches 382; Conservative 0; Mismatches 44; Indels 20; Gaps 4;
XX
QY 66 ACTGAGAGACGAGCTAGTGGATTTCTAGAGCCGATTAAGATCCCTAAGCTTAGCTG 125
Db 1473 ATTAAAGACGAGACGACCTGATGTCTTAGGCTGATTAAGATCCCTAAGCTTAGCTG 1414
QY 126 GAAAGTACGACGCTCACTTTAAACAGCGGGCTTGC AACTTAGCTCAACCTGACCAAT 185
Db 1413 GAAAGTACGACGCTCACTTTAAACAGCGGGCTTGC AACTTAGCTCAACCTGACCAAT 1354
QY 186 C-----AGAGAGCTCACTAAATGCTTAATAGCAAGACAGAGGTAAAGAAATA 236
Db 1353 CAGGTAGTAAAGAGAGCTCACTAAACGCTATTAGCAAAAGAGAGGTAAAGAAATA 1294
QY 237 GCCAATCATCTATTGCTGAGAGACAGACAGAGGAGCAACATCGGGATTAACCCAG 296
Db 1293 GCCAATCATCTATTGCTGAGAGACAGACAGGAGGAGCAACATCGGGATTAACCCAG 1234
QY 297 GCATTGAGCTGGCAACAGACGCCCCCTTTGGGCTCCCTCCCTTTGTATGGAGCTGTT 356
Db 1233 GCATTGAGCTGGCAACAGGCG-TACCTCTCTGGGCTCACTCCCTTTGTATGGAGCTCTG 1175
QY 357 TTCAATGCTATTCTACTCTATTAAATCTTGCACTG--TACTCTTCTGCTCAATGTTCTT 414
Db 1174 T-----TTTCACTCTATTAAATCTTGCACTGCAACTCTTCTGCTCAATGTTCTT 1123
QY 415 ACGGCTGAGCTGAGCTTTTGTCTACCGTCCACGACACTCTGTTTGCACACCGAGACC 474
Db 1122 ATGCTTCAGCTGAGCTTTGCTGCTGACATCCACGACACTCTGTTTGCACACCGAGACC 1063
QY 475 TGCCGCTGACTCCCATCCCTCTGGAT 500
Db 1062 CGTAACTGACTTCCACCCCTCCGGAT 1037
XX
RESULT 13
AAS88392
ID AAS88392 standard; cDNA; 808 BP.
XX
AAS88392;
AC 13-FEB-2002 (first entry)
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #24196.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW Food Supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI, 2001-639362/73.
XX
DR P-PSDB; ABE24205.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 1; SEQ ID No 24196; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 808 BP; 184 A; 234 C; 189 G; 201 T; 0 other;
XX
Query Match 58.5%; Score 292.4; DB 23; Length 808;
Best Local Similarity 85.2%; Pred. No. 8.6e-92;
Matches 380; Conservative 0; Mismatches 46; Indels 20; Gaps 4;
XX
QY 66 ACTGAGAGACGAGCTAGTGGATTTCTAGAGCCGATTAAGATCCCTAAGCTTAGCTG 125
Db 6 ATTAAAGACGAGACGACCTGATGTCTTAGGCTGATTAAGATCCCTAAGCTTAGCTG 65
QY 126 GAAAGTACGACGCTCACTTTAAACAGCGGGCTTGC AACTTAGCTCAACCTGACCAAT 185
Db 66 GAAAGTACGACGCTCACTTTAAACAGCGGGCTTGC AACTTAGCTCAACCTGACCAAT 125
QY 186 C-----AGAGAGCTCACTAAATGCTTAATAGCAAGACAGAGGTAAAGAAATA 236
Db 126 CAGGTAGTAAAGAGAGCTCACTAAACGCTATTAGCAAAAGAGGTAAAGAAATA 185
QY 237 GCCAATCATCTATTGCTGAGAGACAGACAGAGGAGCAACATCGGGATTAACCCAG 296
Db 186 GCCAATCATCTATTGCTGAGAGACAGACAGGAGGAGCAACATCGGGATTAACCCAG 245
QY 297 GCATTGAGCTGGCAACAGACGCCCCCTTTGGGCTCCCTCCCTTTGTATGGAGCTGTT 356
Db 246 GCATTGAGCTGGCAACAGGCG-TACCTCTCTGGGCTCACTCCCTTTGTATGGAGCTCTG 304
QY 357 TTCAATGCTATTCTACTCTATTAAATCTTGCACTG--TACTCTTCTGCTCAATGTTCTT 414
Db 305 T-----TTTCACTCTATTAAATCTTGCACTGCAACTCTTCTGCTCAATGTTCTT 356
QY 415 ACGGCTGAGCTGAGCTTTTGTCTACCGTCCACGACACTCTGTTTGCACACCGAGACC 474
Db 357 ACAGCTTGAAGCTGAGCTTTTGTCTACCGCTCCACGACACTGCTGTTTGTGCGGTACAGACC 416
QY 475 TGCCGCTGACTCCCATCCCTCTGGAT 500
Db 417 CACGCTGACTTCCACCCCTCCGAAT 442
XX
RESULT 14
AAS84189/c
ID AAS84189 standard; cDNA; 1243 BP.
XX
AC AAS84189;
XX

DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #19993.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG20002.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
PS
XX Claim 1; SEQ ID No 19993; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 1243 BP; 285 A; 334 C; 358 G; 266 T; 0 other;
SQ
Query Match 58.2%; Score 290.8; DB 23; Length 1243;
Best Local Similarity 85.0%; Pred. No. 4e-91;
Matches 379; Conservative 0; Mismatches 47; Indels 20; Gaps 4;
QY 66 ACTGAGAGACGAGCTAGTGGATTTCTAGGCGCAGTAAGATCCCTAGCCTAGCTGG 125
DB 1238 ATTGAAGAGACGAGCAGCTGAGTGTCTAGGCTGAGCTAAAGTCCCTAGCCTAGCTGG 1179
QY 126 GAAGGTGACCGCTCCACCTTTAAACAGGGGCTTGCACTTAGCTACACCTGACCAAT 185
DB 1178 GAAGGTGACCTGACCTCACTTTAAACAGGGGCTTGCACTTAGCTACACCTGACCAAT 1119
QY 186 C-----AGAGAGCTCACTAAATGCTAATTAGGCAAGAGGAGGAAAGAAATA 236
DB 1118 CAGGTAGTAAGAGAGCTCACTTAAGCCTATTAGGCAAGAGGAGGAAAGAAATA 1059
QY 237 GCCAATCATCTATTGCTGAGAGCAGCAGAGGAGCAACCAATCGGAGTATAAACCCAG 296

DB 1058 GCCAATCATCTATTGCTGAGAGCAGCAGGAGGAGGAGCAATGACCGGATTTAAACCCAG 999
QY 297 GCATTGAGCTGGCAACAGAGCCCTTTGGGTCCCTTGGCTTGTATGGAGCTGT 356
DB 998 GCATTGAGCTGGCAACAGG-C-TACCTCTCTGAGTCACTTCTGTATGGAGCTCTG 940
QY 357 TTCATGCTATTTCACCTATTAAATCTTGCACTG--CACTCTTGTGTCATGTTCTT 414
DB 939 T-----TTTCACTCTATTAAATCTTGACCTGACACTTTTCTGTGCTCGTGTGTT 888
QY 415 ACGGCTGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTTGCACCAACGAGACC 474
DB 887 AGGCTCTGAGCTGAGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 828
QY 475 TGGCGCTGACTCCCATCCCTCTGAT 500
DB 827 CGCAGCTGACTCTTACCCCTCCGAT 802
RESULT 15
AAK87491
ID AAK87491 standard; DNA; 2629 BP.
XX
XX AAK87491;
XX
XX 07-NOV-2001 (first entry)
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42303.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX MO200157182-A2.
XX
XX 09-AUG-2001.
PF 17-JUN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 15-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.

PR	22-AUG-2000	2000US-0226886
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227109
PR	30-SEP-2000	2000US-0228982
PR	01-SEP-2000	2000US-0228287
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229519
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235835
PR	29-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0235839
PR	29-SEP-2000	2000US-0235839
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239355
PR	13-OCT-2000	2000US-0239357
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	01-NOV-2000	2000US-0244617
PR	01-NOV-2000	2000US-0244618
PR	01-NOV-2000	2000US-0244619
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207

Query Match	57.5%	Score 287.6	DB 22	Length 2629
Best Local Similarity	84.5%	Pred. No. 8.3e-90		
Matches 377	Conservative 0	Mismatches 49	Indels 20	Gaps 4
<p>66 ACTGAGACAGGAGCTAGCTGATTTCTTAGGCCGACTAAGAACTCTTAGCTAGCTGG 125</p> <p>Db 81 ACTGAGACAGGAGCTAGCTGATTTCTTAGGCCGACTAAGAACTCTTAGCTAGCTGG 140</p> <p>126 GAAAGTGACACGCTCCACTTTAAACACGGGGCTTGCACTTAGCTCAACCTGACCAAT 185</p>				

```

Db 141 GAAATTGACCAAGGCTTAAACAGGGGCTTGCAATTAGTCACACCCCAAT 200
QY 186 C-----AGAGGCTCACTAAATGCTAATTAGGCAAGAGGTAAGAAATA 236
Db 201 CAGGTAGTAAAGAGGCTCACTAAATGCTAATTAGGAAACAGAGGTAAGAAATA 260
QY 237 GCCAATCATCTAATTGCTGAGAGCAGACAGAGGGAACAATCGGATATAAACCAG 296
Db 261 GCCAATCATCTAATTGCTGAGAGCAGACAGAGGGAACAATGATCAAGATATAAACCAG 320
QY 297 GCATTGAGCTGGCAAGCAGACCCCTTGGGTCCTTCTTGTATGGAGCTGTT 356
Db 321 GCATTGAGCTGGCAAGCAGACCCCTTGGGTCCTTCTTGTATGGAGCTGTT 379
QY 357 TTCATGCTAATTGCTAATTAAATCTTGAACG--CACCTCTGAGTCATGTTCTT 414
Db 380 T-----TTTCACTAATTAAATCTTGAATGCAACACTTCTGTGATGATGTC 431
QY 415 ACGGCTGAGCTGAGCTTTTGTCTCACCCTCCACCACTGCTGTTGCCACACGAGACC 474
Db 432 ACGGCTGAGCTGAGCTTTTGTCTCACCCTCCACCACTGCTGTTGCCACACGAGACC 491
QY 475 TGCCTGACTCCCATCCCTCTGGAT 500
Db 492 CACAGCTGACTCCCATCCCTCTGGAT 517

```

Search completed: April 17, 2003, 05:06:35
 Job time : 198.426 secs

306 CIGGACACGAGCCUCCCTTGGGICCCCTCCCTTGGAGGAG--CIGTITCAIGC 363

Sequence 3	Appl.1
Sequence 5	Appl.1
Sequence 1	Appl.1
Sequence 1	Appl.1
Sequence 55	Appl.2
Sequence 5	Appl.1
Sequence 5	Appl.1
Sequence 5	Appl.1
Sequence 5	Appl.1
Sequence 5	Appl.1
Sequence 4	Appl.1
Sequence 18	Appl.1
Sequence 2	Appl.1
Sequence 2	Appl.1
Sequence 2	Appl.1
Sequence 2	Appl.1
Sequence 54	Appl.1

```

QY      364  TATTCACTCTATTAACTTGCACACTGCA  :93
          |||||
Db      2900  TATTCACTCTATTAACTTGCACACTGCA  2929
          |||||

```

RESULT 2

```

US-08-686-878A-50
: Sequence 578A-50
: Patent No. 5708157
:
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavoille, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Evans, Cheryl
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: TITLE OF INVENTION: ENCODING THEM
: NUMBER OF SEQUENCES: 71
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/686,878A
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876--551
: INFORMATION FOR SEQ ID NO: 50:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 279 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-08-686-878A-50

```

Query Match	40.0%	Score 200:	LB 1;	Length 279;
Best Local Similarity	88.2%	Pred. No.	4.4e-60;	
Matches 232; Conservative		1; Mismatches	27;	Indels 3; Gaps 2

QY	133	ACAGCTCACACCTTTAAACAGGGGGCTTGCACTTAGCTCACCTACCTACCAATGAGAG	192
	:		
Db	1	RCACATCCACCTTTAAACAAGGGGNTTGCAAANBAGATNACACTTGACCAATCAGAG	60
QY	193	CTCACTAAATGCTATTATAGGCAAAGACAGAGGTAAAGAAATAGCCATCATCTATTGC	252
	:		
Db	61	NTCANTAAATGATNATTNTGGCAAAAACAGAGGTTANAGATAGCCATCATCTATTGC	120
QY	253	CTGAGAGCACAGCAGAGAGGACACAAATCGGGATTAAACCCAGCATTCAGCTGGCAA	312
	:		
Db	121	CTGAGAGCACAGCAGAGAGGACAAATGATCGGGATTAAACCCAGTTTNNAGCGCGCAA	180
QY	313	CAGGACCCCCCTTTGGGGCTTCCTCCCTTTGATGAGGAC--TGTTTTCATGCTATTCCA	370
	:		
Db	181	CGGCA--ACCCCCCTTTGGGTCCTCCCTCTTTGTATGGGACATNTGTTTTCATGCTATTCCA	235
QY	371	CTCTATTAAATCTTGCAACTGCA	393

Db 240 NTNTATTAATNTGCAACTGCA 262

RESULT 3

```

US-08-721-489-4
? Sequence 4, Application US/08721489
? Patent No. 5786465
?
? GENERAL INFORMATION:
? APPLICANT: Jacobs, Kenneth
? APPLICANT: McCoy, John
? APPLICANT: Lavallee, Edward
? APPLICANT: Racie, Lisa
? APPLICANT: Metberg, David
? APPLICANT: Treacy, Maurice
? APPLICANT: Spaulding, Viki
? TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
? TITLE OF INVENTION: ENCODING THEM
? NUMBER OF SEQUENCES: 5
?
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Genetics Institute, Inc.
? STREET: 87 CambridgePark Drive
? CITY: Cambridge
? STATE: Massachusetts
? COUNTRY: U.S.A.
?
? ZIP: 02140
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/721,489
? FILING DATE:
? CLASSIFICATION: 530
?
? ATTORNEY/AGENT INFORMATION:
? NAME: Brown, Scott A.
? REGISTRATION NUMBER: 32,724
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 498-8224
? TELEFAX: (617) 876-5851
?
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 279 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
?
? MOLECULE TYPE: cDNA
?
US-08-721-489-4

```

Query Match	40.0%	Score 200	DB 1	Length 279
Best Local Similarity	88.2%	Pred. NC. 4.8e-60		
Matches 232, Conservative	1	Mismatches 27	Indels 3	Gaps 2

Qy	133	ACAGGTCACTTTAAACAAGGGGGTTGGCAACTTACTCAACCTGACCAATCAGAG	182
Qy	1	RCACATCCACTTTAAACAGGGGGTTGGCAAAAATTAACCTTACCAATCAGAG	60
Qy	193	CTCACTAAATGCTAATTAGGCAAGA CAGAGGTTAAAGAAATAGCCAATCATCTATTGC	252
Db	61	NTCAATTAATTAATTTATGTCAAAAACAGAGGTTAAAGAAATAGCCAATCATCTATTGC	120
Qy	253	CTGAGAGCAACGACGAGGGGCAACATTCGGGATATTAACCCAGGGCATTTGAGGTGGCA	312
Db	121	CTGAGAGCAACGACGAGGGGCAATGATCGGATATTAACCCAGATTGTTAGAGCGGCA	180
Qy	313	CAGCAGCCCCCTTTGGGTCCCTTCCTTTGATGGAGC--TGTTTCAATGCTATTCA	370
Db	181	CGGCA-ACCCCTTTGGGTCCCTTCCTTTGATGGAGCTTTGTTTATGCTATTCA	239
Qy	371	CTCTATTAAATCTTGCAACTGCA	393
Db	240	NTNTATTAATTTTGCAACTGCA	262

Db 1055 CCT 1053

```

|||
RESULT 6
US-08-973-273-1/c
; Sequence 1, Application US/08973273
; Patent No. 6140085
; GENERAL INFORMATION:
; APPLICANT: Dean, Caroline
; APPLICANT: MacKnight, Richard C
; APPLICANT: Bancroft, Ian
; APPLICANT: Lister, Clare K
; TITLE OF INVENTION: Genetic Control of Flowering
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye P.C.
; STREET: 1100 No. 6140085th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,273
; FILING DATE: 01-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01332
; FILING DATE: 03-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9511196.9
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ms Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9763 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; STRAIN: Columbia
; US-08-973-273-1

Query Match 6.2%; Score 30.8; LB 3; Length 9763;
Best Local Similarity 63.5%; Pred. No. 3.5;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 176 CCTGACCATCGAGAGCTCACTAAATGCTAATTAGCGCAAGACGAGTAAAGAAAT 235
Db 4845 CTTGAGAAAGAGACTATATCTCTGATAGTAAAGCAAGAAACGAGGAAAGAAA 4786

QY 236 AGCCATCATCTTAT 249
Db 4785 GGATCAACATCCAT 4772

RESULT 7
US-08-933-750C-90
; Sequence 90, Application US/08933750C
```

```

; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shan, Fuyi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SININOT01
; CLONE: 2184712
; US-08-933-750C-90

Query Match 6.1%; Score 30.6; DB 2; Length 1078;
Best Local Similarity 58.3%; Pred. No. 1.2;
Matches 74; Conservative 0; Mismatches 49; Indels 4; Gaps 1;

QY 288 TAAACCCAGGCAATTCGAGCTGCGCAACAGCAGCCCTTTGGTCCCTTGTATG 347
Db 941 TCAACTGGAAGAGAGATCTTGCTACAGAGAGAGCCCTTGCGCTCCCTTCCTT---TG 996

QY 348 GGAGCTGTTTCATGCTATTTCACTTATTAATCTTGCAAGTCAAGCTCTGTGCAT 407
Db 997 ATAGAGATTAAATGCTTGTTCCTCAATAACTGGGCAAGATGAATCCTAGTGTAT 1056

QY 408 GTTCTT 414
Db 1057 ACTGCCT 1063

RESULT 8
US-09-234-613-90
; Sequence 90, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
```



```

; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Puri
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SININOT01
; CLONE: 2184712
; US-09-234-613-90

Query Match      6.1%; Score 30.6; DB 3; Length 1078;
Best Local Similarity 58.3%; Pred. No. 1.2;
Matches 74; Conservative 0; Mismatches 49; Indels 4; Gaps 1;

QY 288 TAAACCCAGCATTCGAGCTGCGAAGACGACCCCTTTGGCTCCCTTGTATG 347
DB 941 TCAACTGTGAAGAGATCTTGTCAAGAGAGACCTTGGGCTCCCTTCTT---TG 996
QY 348 GGAAGCTGTTTCATGCTATTTCACTATTAATCTTGAACGACGACCTTGGTCCAT 407
DB 997 ATGACGATTATATGCTTCTTGTCCCAATAAACTGGGCGAGATGAATCTAGTCTAT 1056
QY 408 GTTTCCT 414
DB 1057 ACTGCCT 1063

RESULT 9
US-09-739-455-3
; Sequence 3, Application US/09739455
; Patent No. 6413756
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
```

```

; FILE REFERENCE: CL000653
; CURRENT APPLICATION NUMBER: US/09/739,455
; CURRENT FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 11827
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(11827)
; OTHER INFORMATION: n = A,T,C or G
; US-09-739-455-3

Query Match      6.1%; Score 30.6; DB 4; Length 11827;
Best Local Similarity 51.1%; Pred. No. 4.6;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 117 CCTAGCTGGAGAGGTGACACGCTCCACCTTTAAACACGGGGCTTGCACTAGCTCAC 176
DB 10853 CCACGCTGCTGAGAGTGGAACTGCTGCTTTGTTGGCGGCTTTTCTTAATCAGTT 10912
QY 177 CTGACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATA 236
DB 10913 CCTCTTAGATTATTAATACATAAAAAAAATTAAGTTTGAAGAAATAGAGATA 10972
QY 237 GCCAATCATCTATTGCTGANG 257
DB 10973 CAGAAACATGATTTCCACGAG 10993

RESULT 10
US-09-741-154-3/C
; Sequence 3, Application US/09741154
; Patent No. 6437110
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen M. et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001061
; CURRENT APPLICATION NUMBER: US/09/741,154
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 16389
; TYPE: DNA
; ORGANISM: Human
; US-09-741-154-3

Query Match      6.0%; Score 29.8; DB 4; Length 16389;
Best Local Similarity 63.0%; Pred. No. 11;
Matches 46; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 74 ACAAGACTAGCTGATTTCTTCTAGCCGACCTAAGAAATCCCTAAGCTTGGAGGATGA 133
DB 1196 ACAAGACTAATGTGAATGCTGAGTCAATTAATTCCTGATCCAGCTGGAGAGGG 1137
QY 134 CCACGTCACCTT 146
DB 1136 ATGCTCTGAGCTT 1124

RESULT 11
US-08-961-083-199
; Sequence 199, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
; NUMBER OF SEQUENCES: 452
```

```

CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8512
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 2023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-199

Query Match
Best Local Similarity 5.8%; Score 29; DB 3; Length 2023;
Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 203 TGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCTATTGCTGAGAGCAC 262
DB 1711 TGATTTCAGTTTAAGATAGCAGCTAGAGTAAATGTTATGATCATTTGGCTCAGATGAC 1770

QY 263 A 263
DB 1771 A 1771

RESULT 12
US-08-961-527-71
Sequence 71, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
```

```

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-71

Query Match
Best Local Similarity 5.8%; Score 29; DB 4; Length 32768;
Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 203 TGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCTATTGCTGAGAGCAC 262
DB 7947 TGATTTCAGTTTAAGATAGCAGCTAGAGTAAATGTTAGTATGATGCTGCTCAGATGAC 8006

QY 263 A 263
DB 8007 A 8007

RESULT 13
US-09-738-894A-3/c
Sequence 3, Application US/0973894A
Patent No. 6331423
GENERAL INFORMATION:
APPLICANT: GUGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: CL000636
CURRENT APPLICATION NUMBER: US/09/738,894A
CURRENT FILING DATE: 2000-12-18
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 36651
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(36651)
OTHER INFORMATION: n = A,T,C or G
US-09-738-894A-3

Query Match
Best Local Similarity 5.8%; Score 29; DB 4; Length 36651;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 15 TCCTTTCTGGATGAGGCAAGCGCCTGAGATACAGCAATTACTTTCGAACGTGAGGA 74
DB 6891 TCCTACCTGGAACATGAGCCCTGATGGCTGGGGCTACAGCACCATCTTTCGACACATGGA 6832

QY 75 CAGGACTAG 83
DB 6831 AAAGGCTAG 6823

RESULT 14
US-08-703-809-2
Sequence 2, Application US/08703809
Patent No. 5716808
GENERAL INFORMATION:
APPLICANT: Raymond, Christopher K.
TITLE OF INVENTION: GENETIC ENGINEERING OF PICHIA METHANOLICA
```

```

;
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,809
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 96-18
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-703-809-2
;
; Query Match
; Best Local Similarity 49.3%; Score 28.8; DB 1; Length 3386;
; Matches 75; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
;
; QY 307 TGGCAACAGCAGCCCCCTTGGGTCCCTTGTGATGGAGCTGTTTCATGCTAT 366
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 856 TGGCAACAGATAGTATCTATCTGCTTAATTCATCCACTTGGGAACGCTCTCTTTAC 915
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; QY 367 TTCACTATTAATATCTTGCACTGCACCTCTTGTGTCACAGTTTCTTACGGCTCGAGCT 426
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 916 CCCAGATTCTCAAGCTAATATCTGCCCTTGTCTATTTGCTTTCCTGTAACAACG 975
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; QY 427 GAGCTTTGCTACCGTCACCACTGCTGTT 458
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 976 GAGCTTTGCTCCCATCTCTTGTCTTT 1007
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; RESULT 15
; US-08-703-808-2
; Sequence 2, Application US/08703808
; Patent No. 5736383
; GENERAL INFORMATION:
; APPLICANT: Raymond, Christopher K.
; TITLE OF INVENTION: PREPARATION OF PICHIA METHANOLICA AUXOTROPHIC
; TITLE OF INVENTION: MUTANTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,808
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 96-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-703-808-2
;
; Query Match
; Best Local Similarity 49.3%; Score 28.8; DB 1; Length 3386;
; Matches 75; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
;
; QY 307 TGGCAACAGCAGCCCCCTTGGGTCCCTTGTGATGGAGCTGTTTCATGCTAT 366
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 856 TGGCAACAGATAGTATCTGCTTAATTCATCCACTTGGGAACGCTCTCTTTAC 915
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; QY 367 TTCACTATTAATATCTTGCACTGCACCTCTTGTGTCACAGTTTCTTACGGCTCGAGCT 426
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 916 CCCAGATTCTCAAGCTAATATCTGCCCTTGTCTATTTGCTTTCCTGTAACAACG 975
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; QY 427 GAGCTTTGCTACCGTCACCACTGCTGTT 458
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 976 GAGCTTTGCTCCCATCTCTTGTCTTT 1007
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;

```

Search completed: April 17, 2003, 07:37:57
Job time : 73.6205 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:57:20 ; Search time 123.169 Seconds
(without alignments)
4077.979 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500
Perfect score: 500
Sequence: 1 cccctggggcggcctcctt.....tgactccatccctcgtgat 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published_Applications_NA:*
2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
3: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
7: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	348.2	69.6	1329	7 US-08-979-847-108	Sequence 108, App
2	301	60.2	180557	12 US-10-003-806-6	Sequence 6, Appl
3	301	60.2	180557	12 US-10-003-806-6	Sequence 6, Appl
4	284.8	57.0	2946	9 US-10-114-893-134	Sequence 134, App
5	283.8	56.8	2930	10 US-09-902-535-1	Sequence 1, Appl
6	270.4	54.1	635	7 US-08-979-847-102	Sequence 102, App
7	261.4	52.3	410	10 US-09-880-107-538	Sequence 538, App
8	259.2	51.8	541	10 US-09-864-761-8173	Sequence 8173, App
9	232.2	46.4	326014	10 US-09-731-231A-3	Sequence 3, Appl
10	226.2	45.2	1894	10 US-09-864-761-4444	Sequence 4444, App
11	218	43.6	15425	9 US-10-091-504-1654	Sequence 1654, App
12	218	43.6	15425	10 US-09-764-869-1654	Sequence 1654, App
13	204.6	40.9	569	10 US-09-864-761-14951	Sequence 14951, A
14	200	40.0	279	12 US-10-040-916-50	Sequence 50, Appl
15	126	25.2	246	10 US-09-864-761-20462	Sequence 20462, A
16	118.2	23.6	446	10 US-09-811-284-3	Sequence 3, Appl
17	111.6	22.3	475	10 US-09-864-761-895	Sequence 895, App
18	109.4	21.9	3524	10 US-09-972-724-1	Sequence 1, Appl
19	102	20.4	409	10 US-09-864-761-4153	Sequence 4153, App

C 20	101	20.2	559	10	US-09-864-761-7501	Sequence 7501, App
C 21	97.8	19.6	504	10	US-09-864-761-7027	Sequence 7027, App
C 22	93	18.6	440	10	US-09-864-761-3694	Sequence 3694, App
C 23	79.2	15.8	579	10	US-09-864-761-13678	Sequence 13678, A
C 24	50.4	10.1	88	10	US-09-864-761-20907	Sequence 20907, A
C 25	49.8	10.0	525	10	US-09-893-737-31	Sequence 31, Appl
C 26	48.8	9.8	367	10	US-09-864-761-30194	Sequence 30194, A
C 27	37.8	7.6	275	10	US-09-864-761-17675	Sequence 17675, A
C 28	34.8	7.0	1508	9	US-09-735-713A-7	Sequence 7, Appl
C 29	34.6	6.9	79	10	US-09-864-761-31488	Sequence 31488, A
C 30	32.6	6.5	384	10	US-09-783-590-10544	Sequence 10544, A
C 31	32.6	6.5	653	9	US-10-184-644-402	Sequence 402, App
C 32	32.6	6.5	653	9	US-10-184-634-402	Sequence 402, App
C 33	31.8	6.4	468	9	US-09-796-692-4666	Sequence 4666, App
C 34	31.2	6.2	473	10	US-09-864-761-11001	Sequence 11001, A
C 35	31.2	6.2	10514	10	US-09-764-877-3470	Sequence 3470, App
C 36	31.2	6.2	197997	10	US-09-822-246-3	Sequence 3, Appl
C 37	31.1	6.2	407	10	US-09-864-761-20523	Sequence 20523, A
C 38	30.8	6.2	2000	9	US-09-938-842A-2796	Sequence 2796, App
C 39	30.6	6.1	410	10	US-09-867-701-4129	Sequence 4129, App
C 40	30.6	6.1	1078	10	US-09-840-781-90	Sequence 90, Appl
C 41	30.6	6.1	8230	10	US-09-797-908-3	Sequence 3, Appl
C 42	30.6	6.1	8532	10	US-09-817-181-3	Sequence 3, Appl
C 43	30.2	6.0	414	9	US-10-123-155-418	Sequence 418, App
C 44	30.2	6.0	2012	9	US-09-746-783-147	Sequence 147, App
C 45	30.2	6.0	2014	9	US-10-036-041-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-08-979-847-108
Sequence 108, Application US/08979847
Publication No. US00030039664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KONTIRIAN-PRADH, FLORENCE
APPLICANT: JOLIVET-REYNARD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 108:

SEQUENCE CHARACTERISTICS:
LENGTH: 1329 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-108

Query Match 69.6%; Score 348.2; DB 7; Length 1329;
Best Local Similarity 90.2%; Pred. No. 8.4e-112;
Matches 394; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

Qy 66 ACTGAGAGCAGAGCTAGCTGATTTCTAGGCGCACTAAGATCCCTAAGCTAGCTG 125
Db 547 ACTGAGAGCAGAGCTAGCTGATTTCTAGGCTAAGATCCCTAAGCTAGCTG 606
Qy 126 GAAGGTACACGCTCCACTTTAAACAGGGGCTTGCACTTAGCTCACACTGACCAAT 185
Db 607 GAAGGTACCGCATCTCTTAAACATGGGGCTTGCACTTAGCTCACCGGACCAAT 666
Qy 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCAATCAT 245
Db 667 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGCAATAGCAATCAT 726
Qy 246 CTATTGCTGAGAGCAGCAGAGAGGAGACACAAATCGGATATTAACCGGCACTTCGAG 305
Db 727 CTATTGCTGAGAGCAGCAGAGAGGAGACAGGATTTGGATATTAACCGGCACTTCAG 786
Qy 306 CTGGGCAACAGACCGCCCTTTGGGTCCCTTCTTGTATGGAG--CTGTTTCAATGC 363
Db 787 CCAGCAACAGCA-ACCGCCCTTTGGGTCCCTTCTTGTATGGAGCTCTGTTTCACTGC 845
Qy 364 TATTTCACCTTAATTAATCTGCACCTGCTTCTGCTCATCTTTCTTACGGCTCGA 423
Db 846 TATTTCACCTTAATTAATTAATCACTGACCTCTTCTGCTCATCTTTCTTACGGCTCAA 905
Qy 424 GCTGAGCTTTTGTCTGACCGCTCCACCACTGCTTTTCCACCAACCGCAGACTCGCGCTGA 483
Db 906 GCTGAGCTTTTGTGCGCATCCACCACTGCTTTTCCACCAACCGCAGACTCGCGCTGA 965
Qy 484 CTCGCATCCCTCTTGAT 500
Db 966 CTCGCATCCCTCTTGAT 982

RESULT 2
US-10-003-806-6/c
Sequence 6, Application US/10003806
Patent No. US2002011929A1
GENERAL INFORMATION:
APPLICANT: Bishop, Colin E.
APPLICANT: Agoulnik, Alexander I.
APPLICANT: Zhu, Qichao
TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
FILE REFERENCE: P02066US1/10024824
CURRENT APPLICATION NUMBER: US/10/003,806
CURRENT FILING DATE: 2001-11-02
PRIOR APPLICATION NUMBER: 60/245,872
PRIOR FILING DATE: 2000-11-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 180557
TYPE: DNA
ORGANISM: Human
US-10-003-806-6

Query Match 60.2%; Score 301; DB 12; Length 180557;
Best Local Similarity 83.2%; Pred. No. 5.3e-94;
Matches 397; Conservative 0; Mismatches 60; Indels 20; Gaps 4;
Qy 35 AAAGCCTGAGATACAGCAATTAATCTTGCAACTGAGACAGAGCTAGCTGATTTCT 94

Db 59022 AAGGACCTGAGTCAATGGAATATGATCATATGAGAGACAGAGCTAGCTGATTTCT 58963
Qy 95 AGGCGCACTAAGATCCCTAAGCTAGCTGGAAGTGAACCACTCCACTTTAAACAG 154
Db 58962 AGGCGCACTAAGATCCCTAAGCTAGCTGGAAGTGAACCACTCCACTTTAAACAG 58903
Qy 155 GGGCTTGCACTTAGCTCACACCTGACCAATC-----AGAGCTCACTTAAATGC 205
Db 58902 GGGCTTGCACTTAGCTCACACCGCAATCAGATAGTAAAGAGACTCACTTAAATGC 58843
Qy 206 TAATTAGCAAGACAGAGGTAAAGAAATGCCAATCATATTTGCTGAGACACAGC 265
Db 58842 TAATTAGCAAGAAACAGAGGTAAAGAAATGCCAATCATATTTGCTGAGACACAGC 58783
Qy 266 AGAGGAGCAACAATCGGATATTAACCCAGGCAATTCAGCTGAGCAACAGAGCCCT 325
Db 58782 AGAGGAGCAACAATCGGATATTAACCCAGGCAATTCAGGCAACAGC-TACGCTCT 58724
Qy 326 TTGGGTCCCTTCCCTTTGTATGGAGCTGTTTGCATGCTATTTCACTTATTAATCTTG 385
Db 58723 TTGGGTCCCTTCCCTTTGTATGGAGCTGCT-----CTTCACTTATTAATCTTG 58672
Qy 386 CACTGCA-CTCTTCTGCTCATGTTTCTTAACGCTGAGCTGAGCTTTGCTCACCGT 443
Db 58671 CAGCTGCACTCTCTTTGCTCATGTTTCTTAACGCTGAGCTTCTCTGCGCGT 58612
Qy 444 CCACCACTGCTGTTTGCCACACCGCAGACCTGCGCTGACTCCCATCCCTTGAT 500
Db 58611 CCACCACTGCTGTTTGCCGCTGCTGCGACACTGCTGCTGACTTCCATCCGTCAGAT 58555

RESULT 3
US-10-003-806-9/c
Sequence 9, Application US/10003806
Patent No. US2002011929A1
GENERAL INFORMATION:
APPLICANT: Bishop, Colin E.
APPLICANT: Agoulnik, Alexander I.
APPLICANT: Zhu, Qichao
TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
FILE REFERENCE: P02066US1/10024824
CURRENT APPLICATION NUMBER: US/10/003,806
CURRENT FILING DATE: 2001-11-02
PRIOR APPLICATION NUMBER: 60/245,872
PRIOR FILING DATE: 2000-11-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 180557
TYPE: DNA
ORGANISM: Human
US-10-003-806-9

Query Match 60.2%; Score 301; DB 12; Length 180557;
Best Local Similarity 83.2%; Pred. No. 5.3e-94;
Matches 397; Conservative 0; Mismatches 60; Indels 20; Gaps 4;

Qy 35 AAAGCCTGAGATACAGCAATTAATCTTGCAACTGAGACAGAGCTAGCTGATTTCT 94
Db 59022 AAGGACCTGAGTCAATGGAATATGATCATATGAGAGACAGAGCTAGCTGATTTCT 58963
Qy 95 AGGCGCACTAAGATCCCTAAGCTAGCTGGAAGTGAACCACTCCACTTTAAACAG 154
Db 58962 AGGCGCACTAAGATCCCTAAGCTAGCTGGAAGTGAACCACTCCACTTTAAACAG 58903
Qy 155 GGGCTTGCACTTAGCTCACACCTGACCAATC-----AGAGCTCACTTAAATGC 205
Db 58902 GGGCTTGCACTTAGCTCACACCGCAATCAGATAGTAAAGAGACTCACTTAAATGC 58843
Qy 206 TAATTAGCAAGACAGAGGTAAAGAAATGCCAATCATATTTGCTGAGACACAGC 265
Db 58842 TAATTAGCAAGAAACAGAGGTAAAGAAATGCCAATCATATTTGCTGAGACACAGC 58783

```
QY 266 AGGAGGACAACTCGGGATATTAACCCAGGACTTCGAGCTGGCAACAGACGCCCTCT 325
|||
Db 58782 AGGAGGACAACTCGGGATATTAACCCAGGACTTCGAGCTGGCAACAGGCTCT 58724
QY 326 TTGGGTCCTTCCCTTTGATAGGAGCTGTTTTCATGCTATTTCACTCTAATAATCTTG 385
|||
Db 58723 TTGGGTCCTTCCCTTTGATAGGAGGAGCTGT-----CTTCACTCTAATAATCTTG 58672
QY 386 CCACTGCA--CTTCTGTCATGTTTCTTACGAGCTCGAGCTTTTGTCCACGGT 443
|||
Db 58671 CAGTGCACCTCTTTTGGTCTACATTTGTATGATGATTCGAGCTTTTCTCTCGCGGT 58612
QY 444 CCACCACTGCTGTTTGGCCACCGGAGACCTGCGCTGACTCCCATCCCTGGAAT 500
|||
Db 58611 CCACCACTGCTGTTTGGCCGCTGTGCGAGACCTGCTGCTGACTTCATCCATCCGTCGAGAT 58555
```

```
RESULT 4
US-10-114-893-134
; Sequence 134, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallee, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Carlin-Duckett, McKeough
; APPLICANT: Kelleher, Kerry S.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6000-10A
; CURRENT APPLICATION NUMBER: US/10/114,893
; EARLIER FILING DATE: 2002-04-02
; EARLIER FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 134
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-114-893-134
```

```
Query Match 57.0%; Score 284.8; DB 9; Length 2946;
Best Local Similarity 95.5%; Pred. No. 2.6e-89;
Matches 315; Conservative 0; Mismatches 12; Indels 3; Gaps 2;
```

```
QY 66 ACTGAGAGACGAGCTAGCTGATTTCTTACGCGACTAAGAAATCCCTAAGCTTACTGG 125
|||
Db 2601 ACTGAGAGACGAGCTAGCTGATTTCTTACGCGACTAAGAAATCCCTAAGCTTACTGG 2660
QY 126 GAAGGTACACGCTCCACCTTTAAACAGGGGCTTGGCACTTAGCTCACCTGACCAAT 185
|||
Db 2661 GAAGGTACACCATCCACTTTAAACAGGGGCTTGGCACTTAGCTCACCTGACCAAT 2720
QY 186 CAGAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCCATCAT 245
|||
Db 2721 CAGAGAGCTCACTAAATGCTAATTTAGGCAAGAAACAGAGGTAAAGAAATAGCCATCAT 2780
QY 246 CTATTGCTTGAAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTCGAG 305
|||
Db 2781 CTATTGCTTGAAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTCGAG 2840
QY 306 CTGGCAACAGAGCCCTTTGGTCCCTTCTTTATGGGGAG--CTGTTTTCATGC 363
|||
Db 2841 CTGGCAACAGGCA-ACCCCTTTGGTCCCTTCTTTATGGAGGCTCTGTTTTCATGC 2899
QY 364 TATTCACTCTATTAAATCTTGCAACTGCA 393
```

```
Db 2900 TATTCACTCTATTAAATCTTGCAACTGCA 2929
|||
```

```
RESULT 5
US-09-902-535-1
; Sequence 1, Application US/09902535
; Patent No. US20020102530A1
; GENERAL INFORMATION:
; APPLICANT: Keith, Jr., James C.
; APPLICANT: McCoy, John M.
; APPLICANT: Mi, Sha
; TITLE OF INVENTION: Methods and compositions for diagnosing
; TITLE OF INVENTION: and treating preclampsia and gestational trophoblast
; FILE REFERENCE: GIN-6006B4
; CURRENT APPLICATION NUMBER: US/09/902,535
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,657
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (930)...(2546)
US-09-902-535-1
```

```
Query Match 56.8%; Score 283.8; DB 10; Length 2930;
Best Local Similarity 95.4%; Pred. No. 5.8e-89;
Matches 314; Conservative 0; Mismatches 12; Indels 3; Gaps 2;
```

```
QY 66 ACTGAGAGACGAGCTAGCTGATTTCTTACGCGACTAAGAAATCCCTAAGCTTACTGG 125
|||
Db 2603 ACTGAGAGACGAGCTAGCTGATTTCTTACGCGACTAAGAAATCCCTAAGCTTACTGG 2662
QY 126 GAAGGTACACGCTCCACCTTTAAACAGGGGCTTGGCACTTAGCTCACCTGACCAAT 185
|||
Db 2663 GAAGGTACACCATCCACTTTAAACAGGGGCTTGGCACTTAGCTCACCTGACCAAT 2722
QY 186 CAGAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCCATCAT 245
|||
Db 2723 CAGAGAGCTCACTAAATGCTAATTTAGGCAAGAAACAGAGGTAAAGAAATAGCCATCAT 2782
QY 246 CTATTGCTTGAAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTCGAG 305
|||
Db 2783 CTATTGCTTGAAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTCGAG 2842
QY 306 CTGGCAACAGAGCCCTTTGGTCCCTTCTTTATGGGGAG--CTGTTTTCATGC 363
|||
Db 2843 CTGGCAACAGGCA-ACCCCTTTGGTCCCTTCTTTATGGAGGCTCTGTTTTCATGC 2901
QY 364 TATTCACTCTATTAAATCTTGCAACTGC 392
|||
Db 2902 TATTCACTCTATTAAATCTTGCAACTGC 2930
```

```
RESULT 6
US-08-979-847-102
; Sequence 102, Application US/08979847
; Publication No. US2003039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACALA, GLAUCIA
; APPLICANT: KOMRIMAN-PRABHU, FLORENCE
; APPLICANT: JOLIVET-REYNARD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
```

APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-979-847-102
Query Match 54.1%; Score 270.4; 18 7; Length 635;
Best Local Similarity 92.7%; Pred. No. 1.3e-84;
Matches 306; Conservative 0; Mismatches 21; Indels 3; Gaps 2;
QY 66 ACTGAGAGACAGACTAGCTGATTTCTTAGGCCGACATAGATCTTAAAGCTTAGCTGG 125
DB 291 ACTGAGAGACAGACTAGCTGATTTCTTAGGCCGACATAGATCTTAAAGCTTAGCTGG 350
QY 126 GAAGGTGACAGCTCCACTTTAAACAGGGGCTTGCACTTAGCTCAGACCTGACCAAT 185
DB 351 GAAGGTGACAGCTCCACTTTAAACAGGGGCTTGCACTTAGCTCAGACCTGACCAAT 410
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGGAGTAAAGAAATAGCAATCAT 245
DB 411 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAATAGGAGTAAAGAAATAGCAATCAT 470
QY 246 CTATTGCTGAGAGACAGACAGAGGAGCAACATCGGATATAAACCAGGCAATTCAG 305
DB 471 CTATTGCTGAGAGACAGACAGGAGGAGCAAGATCGGATATAAACCAGGCAATTCAG 530
QY 306 CTGGCAACAGACGCCCCCTTTGGGTCCTTCCCTTTGATATGGAG--CTGTTTTCATGC 363
DB 531 CCGGCAACGAGCA-ACCCCCCTTTGGGTCCTTCCCTTTGATATGGGCGCTCTGTTTTCATGC 589
QY 364 TATTTCACCTAATTAATCTTGCAACTGCA 393
DB 590 TATTTCACCTAATTAATCTTGCAACTGCA 619
RESULT 7
US-09-880-107-538/c
Sequence 538, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.

APPLICANT: Scherf, Uwe
TITLE OF INVENTION: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 538
LENGTH: 410
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA250958
US-09-880-107-538
Query Match 52.3%; Score 261.4; DB 10; Length 410;
Best Local Similarity 91.2%; Pred. No. 1.4e-81;
Matches 300; Conservative 0; Mismatches 26; Indels 3; Gaps 2;
QY 66 ACTGAGAGACAGACTAGCTGATTTCTTAGGCCGACATAGATCTTAAAGCTTAGCTGG 125
DB 329 ACTGAGAGACAGACTAGCTGATTTCTTAGGCCGACATAGATCTTAAAGCTTAGCTGG 270
QY 126 GAAGGTGACAGCTCCACTTTAAACAGGGGCTTGCACTTAGCTCAGACCTGACCAAT 185
DB 269 GAAGGTGACAGCTCCACTTTAAACAGGGGCTTGCACTTAGCTCAGACCTGACCAAT 210
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCAT 245
DB 209 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAATAGGAGTAAAGAAATAGCAATCAT 150
QY 246 CTATTGCTGAGAGACAGACAGAGGAGCAACATCGGATATAAACCAGGCAATTCGA 304
DB 149 CTATTGCTGAGAGACAGACAGGAGGAGCAAGATTCGATATAAACCAGGCAATTCGA 90
QY 305 GCTGGCAACAGACGCCCCCTTTGGGTCCTTCCCTTTGATATGGAG--CTGTTTTCATGC 362
DB 89 GCGACGCAACGAGCAACCGCCCTTTGGGTCCTTCCCTTTGATATGGAGCTGTTTTCATGC 30
QY 363 TATTTCACCTAATTAATCTTGCAACTG 391
DB 29 CTATTTCACCTAATTAATCTTGCAACTG 1
RESULT 8
US-09-864-761-8173/c
Sequence 8173, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27


```
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 8173
/ LENGTH: 541
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC016663.2
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.1
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
/ US-09-864-761-8173

Query Match      51.8%; Score 259.2; DB 10; Length 541;
Best Local Similarity 81.6%; Pred. No. 1e-80;
Matches 367; Conservative 0; Mismatches 63; Indels 20; Gaps 5;
```

```
Db      186 ATTCCGGTGGAAATGAGCTTTCTCACCATCCACCATGCTGATGCTGCTGAG 127
Qy      472 AC-CTGCCGCTGAGCTCCCATCCCTGGAT 500
Db      126 ACACCGCTGCTGACTTCCACCCCTCGGAT 97

RESULT 9
US-09-731-231A-3/C
/ Sequence 3, Application US/09731231A
/ Patent No. US2002082189A1
/ GENERAL INFORMATION:
/ APPLICANT: GUEGLER, Karl et al
/ TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
/ TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
/ TITLE OF INVENTION: THEREOF
/ FILE REFERENCE: CL001007
/ CURRENT APPLICATION NUMBER: US/09/731,231A
/ CURRENT FILING DATE: 2000-12-07
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 326014
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(326014)
/ OTHER INFORMATION: n = A,T,C or G
/ US-09-731-231A-3

Query Match      46.4%; Score 232.2; DB 10; Length 326014;
Best Local Similarity 86.7%; Pred. No. 1.2e-69;
Matches 294; Conservative 0; Mismatches 33; Indels 12; Gaps 3;
```

```
Qy      66 ACTGAGAGACAGAGCTAGCTGATTTCTAGGCGGAGCTAGAAATCCCTAAGCTAGCTG 125
Db 170532 ACTGAGAGACAGAGCTAGCTGATTTCTAGGCGGAGCTAGAAATCCCTAAGCTAGCTG 170473
Qy      126 GAAGGTACCAACGCTCCCTTTAAACAGGGGCTTGCACTTAGCTACACCTGACCAAT 185
Db 170472 GAAGGTACGCTCTTCTAAACCCGGGGCTTGCACTTAGCTACACCTGACCAAT 170413
Qy      186 C-----AGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAGAAATA 236
Db 170412 CAGGTAGAAAGAGCTCACTAATAATGTAATAGCTAAGCTAAGAGGTAAGAAATA 170353
Qy      237 GCCAATCATCTATTGCTGAGAGACAGAGGAGCAACAATCGGATATTAACCCAG 296
Db 170352 GCCAATCATCTATTGCTGAGAGACAGAGGAGCAACAATCGGATATTAACCCAG 170293
Qy      297 GCATTGAGCTGACAGACAGACGCCCTTTGGGTCCCTTCCCTTTGATGGAG--CTG 354
Db 170292 GCATTGAGCTGACAGACAGACGCCCTTTGGGTCCCTTCCCTTTGATGGAGCTCTG 170234
Qy      355 TTTTCATGCTATTTCATCTATTAAATCTTGCAACTGCA 393
Db 170233 TTTTCATGCTATTTCATCTATTAAATCTTGCAACTGCA 170195

RESULT 10
US-09-864-761-4444
/ Sequence 4444, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
/ FILE REFERENCE: Aeomica-X-1
```

```

; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine ver. 1.1
; SEQ ID NO 4444
; LENGTH: 1894
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO ACO02346.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
; OTHER INFORMATION: EXPRESSED IN HEBL100, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
; US-09-864-761-4444

Query Match      45.2%; Score 226.2; DB 10; Length 1894;
Best Local Similarity 86.4%; Pred. No. 8.2e-69;
Matches 291; Conservative 0; Mismatches 28; Indels 18; Gaps 3;

```

```

Qy 237 GCCAATCATCTATTGCTGAGACACAGAGGACACAACTCGGATATTAACCCAG 296
Db 1687 GCCAATCATCTGTTGCTGACAGACACAGAGGACAAATGATCGGATATTAACCCAG 1746
Qy 297 GCATTGAGCTGGACAGACAGACCCCTTTGGGTCCTTCCTTTGATGGAGCTGT 356
Db 1747 GCATTGAGCTGGACAGACAGC-TACCCCTTTGGGTCCTTCCTTTGATGGAGCTGT 1805
Qy 357 TTCATGCTATTTCACCTCTATTAAATCTTGCACTGCA 393
Db 1806 T-----CTTCACTCTATTAAATCTTGCACTGCA 1834

RESULT 11
US-10-091-504-1654/c
; Sequence 1654, Application US/10091504
; Publication No. US20030059808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007C1
; CURRENT APPLICATION NUMBER: US/10/091,504
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1654
; LENGTH: 15425
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-091-504-1654

Query Match      43.6%; Score 218; DB 9; Length 15425;
Best Local Similarity 85.5%; Pred. No. 2e-65;
Matches 284; Conservative 0; Mismatches 30; Indels 18; Gaps 3;

```

```

Qy 71 GAGACAGACTGACGTGATTTCTTAGCCGACTAAGAAATCCCTAAGCTGAGGANG 130
Db 1998 GAGACAGAACTAGCTGATTTCTTAGCTCGACTAAGAAATGCTAAGCTGAGGANG 1939
Qy 131 TGACCACTCCACCTTTAAACAGGCGCTTGCACTTAGCTCAGACCTGACCAATC---- 186
Db 1938 TGACTGACATCCATCTTTAAACAGGCGCTTGCACTTAGCTCAGACCTGACCAATCAGT 1879
Qy 187 -----AGAGAGCTCACTAAATGCTAATTAGCAAAAGACAGAGGTAAGAATAGCCA 241
Db 1878 ATTAAGAGAGCTCAGAAATGCTAATTAGCAAAAGACAGAGTAAACAAATAGCCA 1819
Qy 242 TCATCTATTGCTGAGACACAGACAGAGGACAAATCGGATATTAACCCAGCAAT 301
Db 1818 TCATCTATTGCTGAGACACAGAGGACAAATCGGATATTAACCCAGCAAT 1759
Qy 302 CGAGCTGCAACAGAGCCCTTTGGGTCCTTCCTTTGATGGAGCTTTTCAT 361
Db 1758 TGAGCTGCAACAGC-TACCCCTTTGGGTCCTTCCTTTGATGGAGCTGT----- 1704
Qy 362 GCTATTTCATCTATTAAATCTTGCACTGCA 393
Db 1703 -----TTTCACTCTATTAAATCTTGCACTGCA 1676

RESULT 12
US-09-764-869-1654/c
; Sequence 1654, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper

```

NUMBER OF SEQ ID NOS: 2442
 SOFTWARE: Patentlin Ver. 2.0
 SEQ ID NO: 1654
 LENGTH: 15425
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-764-869-1654

Query Match 43.6%; Score 218; DB 10; Length 15425;
 Best Local Similarity 85.5%; Pred. No. 2e-65;
 Matches 284; Conservative 0; Mismatches 30; Indels 18; Gaps 3;

QY 71 GAGACAGACTAGCTGATTTCTTCTAGGCGCACTAAGATCCCTAAGCTTGGAGG 130
 DB 1998 GAGACAGACTAGCTGATTTCTTCTAGGCGCACTAAGATCCCTAAGCTTGGAGG 1939
 QY 131 TGACCACTGCTTAAACACGGGGCTTGCACTTACCTCAACCTGACCAATC---- 186
 DB 1938 TGACTGATCATCTTTAAACATGGGGCTTGCACTTACCTCAACCTGACCAATCAGGT 1879
 QY 187 -----AGAGAGCTCACTAATTAATTAAGCAAGAGAGTAAAGAAATAGCCAA 241
 DB 1878 ATTAAGAGAGCTCACTAATTAATTAAGCAAGAGAGTAAAGAAATAGCCAA 1819
 QY 242 TCATCTATTGCTGAGACACAGAGAGGACAACTCGGATATATAACCCAGGCAAT 301
 DB 1818 TCATCTATTGCTGAGACACAGAGAGGACAACTCGGATATATAACCCAGGCAAT 1759
 QY 302 CGAGCTGGCAACAGAGCCCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTTCAT 361
 DB 1758 TGAGCTGGCAACAGC-TACCTCTTTGGGTCCTTCCCTTTGATGGAGCTGTT---- 1704
 QY 362 GCTATTCACTTAAATCTTGAACCTGCA 393
 DB 1703 -----TTTCACTTAAATCTTGAACCTGCA 1676

RESULT 13

US-09-864-761-14951/c
 Sequence 14951 Application US/09864761
 Patent No. US2002048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: Aeomica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 14951
 LENGTH: 569
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AP000233.1
 OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 9.3
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
 US-09-864-761-14951

Query Match 40.9%; Score 204.6; DB 10; Length 569;
 Best Local Similarity 79.4%; Pred. No. 1.6e-61;
 Matches 336; Conservative 0; Mismatches 64; Indels 23; Gaps 7;

QY 66 ACTGAGAGAGAGCTAGCTGATTTCTTCTAGGCGCACTAAGATCCCTAAGCTTGG 125
 DB 415 ACCGAGAGAGAGCTAGCTGATTTCTTCTAGGCGCACTAAGATCCCTAAGCTTGG 356
 QY 126 G-AAGTGACCACTGCTTAAACACGGGGCTTGCACTTACCTGACCTGACCA 184
 DB 355 GAAAGTGACCACTGCTTAAACACGGGGCTTGCACTTACCTGACCTGACCA 296
 QY 185 TC-----AGAGAGCTCACTTAAATGCTAATTAGGC-AAAGACAGAGGTTAAAGAA 234
 DB 295 TCAGGTAGTAAAGAGGGGTTTACGTAATACAAATTAGGCTTAAAGAGGAGTAAAGAAA 236
 QY 235 TAG-CCAATCATCTATGCTGAGAGACACAGAGGAGGACAACTCGGATAT-AAAC 292
 DB 235 TAGTCAATCATCTATGCTGAGAGACACAGAGGAGGACAACTCGGATATAAAC 176
 QY 293 CCAAGCATTCGAGCTGCAACAGAGCCCCCTTTGGGTCCTTCCCTTTGATGGAGC 352
 DB 175 CCAAGCATTCGAGCTGCAACAGAGCCCCCTTTGGGTCCTTCCCTTTGATGGAGC 116
 QY 353 TGTTCATGCTATTTCACTCTAATTAATCTTGAACCTG--CACTCTTGGTTCATGTT 410
 DB 115 TGTGT-----TTTCACTCTGTTAAATCTTGAACCTGTAACCTCTGTCAGTGT 64
 QY 411 TCTTACGCTGAGCTGAGCTTTTGTCTCAACCTGCACTGCTGTTTGCACACGCA 470
 DB 63 TGTTCGAGCTGAGCTGAGCTTTTGTCTCAACCTGCACTGCTGTTTGCACACGCA 4
 QY 471 GAC 473
 DB 3 GAC 1

RESULT 14

US-10-040-916-50
 Sequence 50 Application US/10040916
 Patent No. US20020146769A1
 GENERAL INFORMATION:
 APPLICANT: Jacobs, Kenneth
 APPLICANT: McCoy, John
 APPLICANT: Lavallie, Edward

Racie, Lisa
Merberg, David
Treacy, Maurice
Evans, Cheryl
Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040,916
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/887,029
FILING DATE: 07-FEB-1997
APPLICATION NUMBER: 08/586,878
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-040-916-50
Query Match 40.0%; Score 200; DB 12; Length 279;
Best Local Similarity 88.2%; Pred. No. 4.6e-60;
Matches 232; Conservative 1; Mismatches 27; Indels 3; Gaps 2;
QY 133 ACCAGCTCCACCTTTAAACACGGGCTTGCACTTGCATCAGCTGACCATCAGAGNG 192
DB 1 RCCACATCCACTTTAAACACGGGCTTGCACTTGCATCAGCTGACCATCAGAGNG 60
QY 193 CTCCTAAATATCTTAATTAGGCAAGAGGATTAATTAATGCACTTCTATTTC 252
DB 61 NTCANTAAATATGATTAATTGGCAAAAACGAGATTAATTAATGCACTTCTATTTC 120
QY 233 CTGAAGACACAGAGAGGAGCAACAATGGGATATAATCCAGGATTCGAGCTGGCAA 312
DB 121 CTGAAGACACAGAGAGGAGCAACAATGGGATATAATCCAGGATTCGAGCTGGCAA 180
QY 313 CAGGAGCCCCCTTTGGGCTCTCCCTTTGATGGAGC--TGTTTCATCTATTTC 370
DB 161 CGGCA-ACCCCTTTGGGCTCTCCCTTTGATGGAGCCTTGTTCATCTATTTC 239
QY 371 CTCTATTAAATCTTGCACTGCA 393
DB 240 NTMTATTAAATTTGCAACTGCA 262

Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 20462
LENGTH: 246
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010951.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
OTHER INFORMATION: NT HIT: AB026898.1, EVALUATE 4.00e-57
OTHER INFORMATION: EST HUMAN HIT: A1492055.1, EVALUATE 3.00e-49
OTHER INFORMATION: SWISSPROT HIT: Q02279, EVALUATE 7.30e-01
US-09-864-761-20462
Query Match 25.2%; Score 126; DB 10; Length 246;
Best Local Similarity 85.6%; Pred. No. 4.5e-34;
Matches 178; Conservative 0; Mismatches 20; Indels 10; Gaps 3;

```
QY 246 CTATTGCTGAGACGACAGAGAGGACAACTCGGATATATAACCCAGGCATTGAG 305
Db 246 CTGTTGCCCTGAGACACAGCGGAGGACAAATATATCAGATATATAACCAAGCATTCGAG 187
QY 306 CTGACACAGACAGCCCCCTTTGGGTCCTTCCCTTTGTATGGAGCTGTTTCATGCTA 365
Db 186 CTGCAACGGTA-ACCCCCTTTGGGTCCCTTGTATGGAGCTCTAT----- 136
QY 366 TTTCACTCTATTAAATCTTGCAACTGCACCTCTCTGGTCCATGTTTCTTAAGGCTCGAGC 425
Db 135 CTTCACTCTATTAAATCTTGCAACTGTACTCTTGGTCCGTGTTGTAC-GCTTGAGC 77
QY 426 TGAAGTTTGTCTCACCGTCCACCACTGC 453
Db 76 TGAAGTTTGTCTCACCGTCCACCACTGC 49
```

Search completed: April 17, 2003, 07:46:40
Job time : 301.169 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:30:25 ; Search time 1266.31 seconds
(without alignments)
6394.748 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Perfect score: 500

Sequence: 1 ccctgggggggggtcccttc.....tgacctccatccctcgat 500

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

```
Database :
EST:
1: em_estda:*
2: em_esthum:*
3: em_estnu:*
4: em_estnu:*
5: em_estnu:*
6: em_estnu:*
7: em_estnu:*
8: em_estnu:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	336	67.2	727	17	AG029908 Pan trogl
2	332.2	64.4	683	17	AG102951 Pan trogl
3	285.4	57.1	440	12	BE732673 BE732673
4	283.8	56.8	415	9	AI128496 AI128496
5	283.2	56.6	436	9	AI128526 AI128526
6	282.2	56.4	494	9	AA781423 AA781423

C	7	280	56.0	342	9	AA860368	AA860368 aj59c05.s
C	8	277.2	55.4	921	17	BH149565	BH149565 ENT048TR
C	9	273	54.6	422	14	NS3177	NS3177 yv6h11.s1
C	10	271.8	54.4	388	14	H01325	H01325 y199e01.s1
C	11	271.2	54.2	385	14	NS5091	NS5091 yv43e03.s1
C	12	270.8	54.2	342	14	T47345	T47345 yb10h02.s1
C	13	270.6	54.1	771	13	BI087886	BI087886 602852690
C	14	267.4	53.5	522	10	AW971553	AW971553 EST183642
C	15	265	53.0	363	14	T69704	T69704 yd13a03.s1
C	16	263.6	52.7	328	9	AA729556	AA729556 nx58c05.s
C	17	261.4	52.3	410	9	AA250958	AA250958 z807d10.s
C	18	261	52.2	443	9	AA837267	AA837267 od26d10.s
C	19	260.4	52.1	681	10	AV722664	AV722664 AV722664
C	20	259.6	51.9	425	9	AI570707	AI570707 tm79G09.x
C	21	259.6	51.9	433	9	AI379210	AI379210 cd01g11.x
C	22	259.6	51.9	490	9	AI598135	AI598135 tm14a10.x
C	23	258.8	51.8	681	10	AV721910	AV721910 AV721910
C	24	258.2	51.6	424	14	R27412	R27412 yb46d11.s1
C	25	257.6	51.5	458	14	R76086	R76086 y171b03.s1
C	26	256.2	51.2	404	14	R27389	R27389 yb46a09.s1
C	27	253.4	50.7	893	14	BQ437925	BQ437925 AGENCOURT
C	28	253.2	50.6	446	9	AI393478	AI393478 tg45G04.x
C	29	250.8	50.2	611	17	AQ381711	AQ381711 RPL11-16
C	30	246.6	49.3	332	9	AI797629	AI797629 w82a08.x
C	31	245.4	49.1	701	17	AG126669	AG126669 Pan trogl
C	32	243.8	48.8	438	14	R77278	R77278 y175d06.s1
C	33	243.6	48.7	722	17	AG104643	AG104643 Pan trogl
C	34	243.4	48.7	396	9	AA814939	AA814939 oc07d02.s
C	35	239	47.8	722	17	AG049481	AG049481 Pan trogl
C	36	238.6	47.7	326	14	D29167	D29167 HUNM203 Hu
C	37	238.6	47.7	463	14	R68685	R68685 y114G06.s1
C	38	237.6	47.5	447	12	BF919416	BF919416 QV0-WT015
C	39	236.8	47.4	683	17	AG134524	AG134524 Pan trogl
C	40	234.8	47.0	330	9	AA774109	AA774109 ac36d03.s
C	41	233.6	46.7	470	9	AI791155	AI791155 ab52f07.x
C	42	233.6	46.7	490	12	BF919425	BF919425 QV0-WT015
C	43	232.6	46.5	440	9	AI003607	AI003607 zE99a12.s
C	44	232.6	46.5	471	9	AA709471	AA709471 zE91h06.s
C	45	232.4	46.5	757	17	AG121490	AG121490 Pan trogl

ALIGNMENTS

RESULT 1
AG029908 727 bp DNA linear GSS 01-NOV-2001
LOCUS Pan troglodytes DNA, clone: PTB-002C04.R, genomic survey sequence.
DEFINITION AG029908
ACCESSION AG029908.1 GI:16556781
VERSION
KEYWORDS
SOURCE
ORGANISM
BAC library clone: PTB-002C04.R.
Pan troglodytes
Pan troglodytes
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of library PTB
Unpublished
2 (bases 1 to 727)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-chou, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpses@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
COMMENT

PRIMERS

Sequencing: M13Rev
LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

source

1..727

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-002C04.R"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC library"

BASE COUNT 170 a 215 c 175 g 166 t 1 others

ORIGIN

Query Match 67.2%; Score 336; EB 17; Length 727;

Best Local Similarity 87.2%; Pred. No. 3,6e-99;

Matches 380; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

QY 61 TTGCACTGAGAGACGAGCTAGCTGATTTCTGAGCCGATAGAACTCCTAAGCCTA 120

DB 270 TTCTCATGAGAGACGAGCTAGCTGATTTCTGAGCCGATAGAACTCCTAAGCCTA 329

QY 121 GCTGGGAAGTGACGACGCTTAAACAGCGGGTTGCACTTACCTCACAACCTGA 180

DB 330 GCTGGGAAGTGACGACGCTTAAACAGCGGGTTGCACTTACCTCACAACCTGA 389

QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCA 240

DB 390 CCTATCGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCA 449

QY 241 ATCATCTATGCTGAGAGACGAGGAGGAGCAATCGGAGATTAACCCAGGACAT 300

DB 450 ATCATCTATGCTGAGAGACGAGGAGGAGCAATCGGAGATTAACCCAGGACAT 509

QY 301 TCGAGCTGGCAACAGAGCCCTTTGGGCTCCCTTGTATGAGAGCTGTTTCA 360

DB 510 TCAAGCCAGCAACGAGCAACCCCTTTGGGCTCCCTTGTATGAGAGCTGTTTCA 569

QY 361 TCGATTTCTCTATTAATCTTGCACTGCACTCTTGTATGAGAGCTGTTTCA 420

DB 570 CTCTATTTCTCTATTAATCTTGCACTGCACTCTTGTATGAGAGCTGTTTCA 629

QY 421 CGAGCTGAGCTTTGCTCACCGTCCACCACTGCTGT-TGCCAACACCGCAGACCTGCCG 479

DB 630 CGAGCTGAGCTTTGCTCACCGTCCACCACTGCTGT-TGCCAACACCGCAGACCTGCCG 689

QY 480 CTGACTCCCATCCCTC 495

DB 690 TTGACTTCATCCCTC 705

RESULT 2

AG102951 683 bp DVA linear GSS 03-NOV-2001

LOCUS DEFINITION Pan troglodytes DNA, clone: PTB-106G16.F, genomic survey sequence.

ACCESSION AG102951.1 GI:16723468

VERSION AG102951.1 GI:16723468

KEYWORDS Pan troglodytes male lymphoblast DNA, clone: PTB Chimpanzee Male

SOURCE Pan troglodytes

ORGANISM Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Tokoki, Y., Watanabe, H. and Sakaki, Y.

BAC end sequences of library PTB

Unpublished

REFERENCE 2 (bases 1 to 683)

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Unpublished

REFERENCE 2 (bases 1 to 683)

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

TITLE

Direct Submission

Submitted (02-AUG-2001) Ageo Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: chimbes@gsr.riken.go.jp, URL: http://hsp.gsc.riken.go.jp/)

Tel: 81-45-503-9111, Fax: 81-45-503-9170

Clones are derived from the chimpanzee BAC library PTB This BAC end

clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

source

1..683

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-106G16.F"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC library"

BASE COUNT 162 a 186 c 167 g 167 t 1 others

ORIGIN

Query Match 64.4%; Score 322.2; DB 17; Length 683;

Best Local Similarity 87.4%; Pred. No. 1.2e-94;

Matches 387; Conservative 0; Mismatches 53; Indels 3; Gaps 3;

QY 59 TCTTGAAGTGAAGAGACGAGCTAGCTGATTTCTTACGCCAGTAAAGATCCCTAAGCC 118

DB 101 TCTTGAAGTGAAGAGACGAGCTAGCTGATTTCTTACGCCAGTAAAGATCCCTAAGCC 160

QY 119 TAGCTGGAGAGTGAACAGCTTCACTTTAAACAGCGGCTTGCACTTACCTCAGACT 178

DB 161 TAGCTGGAGAGTGAACAGCTTCACTTTAAACAGCGGCTTGCACTTACCTCAGACT 220

QY 179 GACCAATAGAGAGTCACTAAATGCTAATTAGGCAAAAGACAGAGGTTAAAGAAATAGC 238

DB 221 AACCAATAGAGAGTCACTAAATGCTAATTAGGCAAAAGACAGAGGTTAAAGAAATAGC 279

QY 239 CAATCATCTATTTGCTGAGACGACAGAGGAGCAACAATCGGATATTAACCCAGGC 298

DB 280 CAATCATCTATTTGCTGAGACGACAGAGGAGCAACAATCGGATATTAACCCAGGC 339

QY 299 ATTGAGCTGGGAACAGAGCCCTTTGGGCTCCCTTGTATGAGAGCTGTTT 358

DB 340 ATTGAGCTGGGAACAGAGCCCTTTGGGCTCCCTTGTATGAGAGCTGTTT 398

QY 359 CATGCTATTTCACTATTAATCTTGCACTGCACTCTTGTGATGTTTCTTACG 418

DB 399 CACTTATTTCACTATTAATCTTGCACTGCACTCTTGTGATGTTTCTTACG 458

QY 419 CTGAGCTGAGCTTTTGTCTACCGTCCACCACTGCTG-TTTCACACGACCGCAGACTGC 477

DB 459 CTGAGCTGAGCTTTTGTCTACCGTCCACCACTGCTG-TTTCACACGACCGCAGACTGC 518

QY 478 CGCTGACTCCCATCCCTCTGAT 500

DB 519 CGCTGACTCCCATCCCTCTGAT 541

RESULT 3

BE732673

LOCUS

DEFINITION 601571305F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3925728 5',

mrna sequence.

ACCESSION BE732673

VERSION BE732673.1 GI:10146665

KEYWORDS

SOURCE human.

ACCESSION	AI128526	GI:3597040
VERSION	AI128526.1	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 436)	
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
COMMENT	Tumor Gene Index	
	Unpublished (1997)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgapbs-r@mail.nih.gov	
	This clone is available royally-free through LINL ; contact the	
	IMAGE Consortium (info@image.llnl.gov) for further information.	
	Insert length: 720	
	Std Error: 0.00	
	Seq primer: -40m3 fwd. ET from Amersham	
	High quality sequence stop: 428.	
FEATURES	Location/Qualifiers	
SOURCE	1..436	

BASE COUNT	95 a	96 c	106 g	139 t
ORIGIN				
Query Match		56.6%	Score 283.2;	DB 9; Length 436;
Best Local Similarity		95.2%;	Pred. No. 7e-12;	
Matches 314; Conservative		0;	Mismatches 213;	Indels 3; Gaps 2;
QY	66	ACTGAGAGACAGACCTAAGCTGATTTCCCTAGAGCCGACCTAAGATTCCTTAAGCTTACTGG	125	
Db	351	ACTGAAAGACAGAGCTAGCTGAGATTTCCCTAGGCTGACTAAGATTCCTTAAGCTTACTGG	292	
QY	126	GAAGGTGACACCGTCCACCTTTTAAACACGCGGCTTGACACTTAGTCTACACCTGACCAAT	185	
Db	291	GAAAGTGACCACTCCACCTTTTAAACCGGGCTTGCACTTAGTCTACACCTGACCAAT	232	
QY	186	CAGAGAGCTCACTAAATGCTAATTATAGGCAAAGACAGAGATTAAGAAATAGCCAAATCAT	245	
Db	231	CAGAGAGCTCACTAAATATCTAATTATAGGCAAAGACAGAGATTAAGAAATAGCCAAATCAT	172	
QY	246	CAATTGCCCTGAGAGCACAGACGAGGAGGACAAACATGGATATTAATACCAGAGCATTCGAG	305	
Db	171	CAATTGCCCTGAGAGCACAGACGAGGAGGACATGATGGATATTAATACCAGATTTTCGAG	112	
QY	306	CTGGCAACAGACAGCCCCCTTTGGTGGCTTCCCTTGTATGAGGAA--CTGTTTTCATGC	363	
Db	111	CCGGCAACGGCA--ACCCCTTTGGGTCCCTCCCTTTGTGATGGAGACTCGTTTTTCAATGC	53	
QY	364	TATTTCACTCTATTAATATCTTGCACTGCA	393	
Db	52	TATTTCACTTATTAATATCTTGCACTGCA	23	
RESULT 6				
LOCUS	AA781423	494 bp	mRNA	linear
DEFINITION	aJ26C03.sl Soares testis NHT Homo sapiens cDNA Clone 1391428 3'			EST 31-DEC-1996

```

similar to contains PRR7.ct1 PRR7 repetitive element ; , mRNA
sequence.
ACCESSION      AA781423
VERSION        AA781423.1  GI:2840754
KEYWORDS
SOURCE
ORGANISM       human.
                Homo sapiens
                Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 494)
                NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
                National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
JOURNAL        Unpublished (1997)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-remail.nih.gov
                cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
                , Ph.D.
                cDNA Library Arrayed by: Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                www-bio.llnl.gov/bdtp/image/image.html
                Insert Length: 1645      Std Error: 0.00
                Seq primer: -40ml3 fwd, EF from AmerSham
                High quality sequence stop: 475.
                Location/Qualifiers
FEATURES
                1..494
SOURCE

```

/note="vector: pT7SD-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories , Inc., and primed with a Not I - oligo(dT) primer [5' TGTCACCAATCTGAAAGTGAGCGGCCGCCCATTTTGTATTGTTT TTTT 3'] . Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cote's and was constructed by Bento Soares and M. Fatima Bonaldo. "

	BASE COUNT	ORIGIN
	108 a	111 c 133 g 141 t 1 others
Query Match	56.4%;	Score 282.2; DB 9; Length 494;
Best Local Similarity	94.8%;	Pred. No. 1.6e-81;
Matches 313; Conservative	0; Mismatches 14; Indels 3; Gaps 2;	
Dy	66	ACTGAGAGCAGGAAGTAAGTCTGGATTTCCCTAAGCCGCAATAAATCCCTAAGCCTTAGCTGG 125
Dd	334	ACTGAGAGACAGGACTACTGATGATTTCTTAGGCTGACTAAGAATCCCTAAGCCTTAGCTGG 275
Dy	126	GAAAGTGAACAAGTCACACCTTTAAAACAGGGGGCTTGCACCTTAGCTCACACTGACCAT 185
Dd	274	GAAAGTGAACAATCCACTCTTTAAAACAAGGGGCTTGCACCTTAGCTCACACTGACCAT 215
Dy	186	CAGAGAGCTCACTAAATGCTTAATTAGCCAAAGACAGAGGTAAGAATATGCCCAATAT 245
Dd	214	CAGAGAGCTCACTAAATGCTTAAGTACAGCAAAGACAGAGGTTAAAGAAATATGCCCAATAT 155
Dy	246	CTATTGCTCTGAGAGACAGCAGAGGGGACAACATCGGATATATAAACCCAGAGCATTTGAG 305
Dd	154	CTATTGCTCTGAGAGACAGCAGAGGGGACAATGATCGGATATATAAACCCAGAGCATTTGAG 95
Dy	306	CTGGCAACAGCAGCCCCCTTTGGTTCCTTCCCTTTGTATGGAG--CTGTTTTCATGC 363
Dd	94	CCGGCAAGGGCA-AACCCCTTTGGGTCCCCTTCCTTTGTGGGAGCTGTGTTTCATGC 36
Dy	364	TATTTCACTTAATTAATCTTGGCACTGCA 393

Db 35 TATTCACTCTATTAAATCTTGCAACTGCA 6

RESULT 7
AA860368/c
LOCUS 342 bp mRNA linear EST 31-DEC-1998
DEFINITION aj59c05.s1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1394600
3' similar to contains PTR7.c1 PTR5 repetitive element ;, mRNA
sequence.

ACCESSION AA860368
VERSION AA860368
KEYWORDS EST.
SOURCE AA860368.1 GI:2954363
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 342)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/MLN at:
www-bio.llnl.gov/dbfp/image/image.html
Insert Length: 1305
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 319.

FEATURES
source
1..342
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394600"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGACATGAGGAGCGGCCCAATTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cos, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 75 a 78 c 82 g 107 t

ORIGIN

Query Match 56.0%; Score 280; DB 9; Length 342;
Best Local Similarity 94.5%; Pred. No. 7.1e-81;
Matches 312; Conservative 0; Mismatches 15; Indels 3; Gaps 2;

Db 66 ACTGAGAGACGAGACTAGCTGATTTCTAGCCGACTAAAGATCCCTAAGCTTACTGG 125
338 ACTAGAGACGAGACTAGCTGATTTCTAGCCGACTAAAGATCCCTAAGCTTACTGG 279

Db 126 GAAGTGACCAAGTCCACTTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 185
278 GAAGTGACCAAGTCCACTTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 219

Db 186 CAGAGAGCTCACTAAATGCTATTGAGCAAGACAGAGGTAAAGAAATAGCCATCAT 245
218 CAGAGAGCTCACTAAATGCTATTGAGCAAGACAGAGGTAAAGAAATAGCCATCAT 159

Db 246 CTATTGCTGAGAGACGAGGAGGACAAACATCGGATATTAAACCCAGCATTTGAG 305
158 CTATTGCTGAGAGACGAGGAGGACAAATGATCGGGATATTAAACCCAGCATTTGAG 99

Qy 306 CTGGCAACAGCAGCCCCCTTGGGTCCTTCCCTTGTATGGAG--CTGTTTTCATGC 363
Db 98 GCGGCAACGGCA-AACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTATGC 40

Qy 364 TATTCACTCTATTAAATCTTGCAACTGCA 393
Db 39 TATTCACTCTATTAAATCTTGCAACTGCA 10

RESULT 8
BH149565/c
LOCUS 921 bp DNA linear GSS 27-AUG-2001
DEFINITION ENTQ648TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.

ACCESSION BH149565
VERSION BH149565.1 GI:15310303
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 921)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HML:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HML:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 40
High quality sequence stop: 567.

FEATURES
source
1..921
/organism="Entamoeba histolytica"
/strain="HML:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt, Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999). "

BASE COUNT 242 a 216 c 226 g 237 t

ORIGIN

Query Match 55.4%; Score 277.2; DB 17; Length 921;
Best Local Similarity 84.5%; Pred. No. 9.1e-80;
Matches 377; Conservative 0; Mismatches 48; Indels 21; Gaps 5;

Db 66 ACTGAGAGACGAGACTAGCTGATTTCTAGCCGACTAAAGATCCCTAAGCTTACTGG 125
651 AGTGAAGAGACGAGACTAGCTGATTTCTAGCCGACTAAAGATCCCTAAGCTTACTGG 592

Db 126 GAAGTGACCAAGTCCACTTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 185
591 GAAGTGACCAAGTCCACTTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 532

BASE COUNT	95 a	95 c	111 g	120 t	1 others
ORIGIN					
Query Match	54.6%	Score 273;	DB 14;	Length 422;	
Best Local Similarity	93.3%	Pred. No. 1.6e-78;			
Matches 308;	Conservative	0;	Mismatches 16;	Indels 6;	Gaps 2;
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonalito."					
QY	66	ACTGAGAGACAGGACTGAGTGGATTTCCCTAGAGCCGACTAAGATCCCTAAGCCTAGCTGG	125		
Db	327	ACTGAGAGACAGGACTGAGTGGATTTCCCTAGAGCTGACATGAATCCCTAAGCCTAGCTGG	268		
QY	126	GAAAGTGACCAACGTCACCTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAAT	185		
Db	267	GAAAGTGACCAACATCCACCTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAAT	208		
QY	186	CAGAGAGCTCATTAATGCTTAATTAGCCAAAGACAGAGTTAAAGAAATGCGCAATAT	245		
Db	207	CAGAGAGCTCATTAATGCTTAATTAGCCAAAGACAGAGTTAAAGAAATGCGCAATAT	148		
QY	246	CTATTGCTTGAGACACAGCAGAGGAGCAACAATCGGATATTAACCCAGGCAATTCGAG	305		
Db	147	CTATTGCTTGAGACACAGCAGAGGAGCAACAATCGGATATTAACCCAGGCTTCGAG	88		
QY	306	CTGGCAACAGCAGCCGCCCTTTGGGTCCTTCCTTTGTATGGAG--CTGTTTTCATGC	363		
Db	87	CGG----CAACGGCAACCCCTTTGGTCCCTTCCTTTGTATGGAGCTGTGTTTCATGC	32		
QY	364	TATTTCACTCTATTAAATCTTGCAACTGCA	393		
Db	31	TATTTCACTCTATTAAATCTTGCAACTGCA	2		
RESULT 10					
H01325/c					
LOCUS	H01325	388 bp	mRNA	linear	EST 19-JUN-1995
DEFINITION	y199e01.a1 Soares placenta NB2HP Homo sapiens cDNA clone				
IMAGE:	147384 3', mRNA sequence.				
ACCESSION	H01325				
VERSION	H01325.1	GI:864258			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
REFERENCE	1 (bases 1 to 388)				
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Matra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaekis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.				
TITLE	The Washu-Merck EST project				
JOURNAL	Unpublished (1995)				
COMMENT	Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: eec@watson.wustl.edu Insert Size: 790 High quality sequence stops: 346 Source: IMAGE Consortium, LNLU This clone is available royalty-free through LNLU; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 790 Std Error: 0.00 Seq primer: Promega -21ml3 High quality sequence stop: 346. Location/Qualifiers 1..388 /organism="Homo sapiens"				
FEATURES					
SOURCE					

/db_xref="GDB:559031"
/db_xref="taxon:9606"
/clone="IMAGE:147384"
/clone_lib="Soares placenta Mb2HP"
/sex="Female"
/dev_stage="Placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', ACTGGAAGAATTCGGCGGCGGAGGATTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 90 a 90 c 96 g 110 t 2 others
ORIGIN

Query Match 54.4%; Score 271.8; DB 14; Length 388;
Best Local Similarity 93.8%; Pred. No. 3.8e-78;
Matches 305; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

QY 66 ACTGAGAGACGAGACTGCTGATTTCTTAGCGCAGCTAAGATCCCTAGCCTGCTG 125
DB 329 ACTGAGAGACGAGACTGCTGATTTCTTAGCGCAGCTAAGATCCCTAGCCTGCTG 270

QY 126 GAAGGTGACGACGTCACCTTTAAACAGGGGGCTTGCACTAGCTCACCTGACCAAT 185
DB 269 GAAGGTGACGACGTCACCTTTAAACAGGGGGCTTGCACTAGCTCACCTGACCAAT 210

QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGAGGTAAAGAAATGCCAATCAT 245
DB 209 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCAT 150

QY 246 CTATTGCTTGAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 305
DB 149 CTATTGCTTGAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 90

QY 306 CTGCAACAGAGAGGCCCCCTTTGGTGCTCCCTTTGATAGGAG--CTGTTTTCATGC 363
DB 89 GCGGCAACGGCA-AGCCCCCTTTGGTGCTCCCTTTGATAGGAGCTCTGTTTTCATGC 31

QY 364 TATTCACTCTATTAAATCTTGCA 388
DB 30 TATTCACTCTATTAAATCTTGCA 6

RESULT 11 385 bp mRNA linear EST 28-JAN-1997
N55091/c
LOCUS y743803.81 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:245500 3', mRNA sequence.

ACCESSION N55091
VERSION N55091.1 GI:1197970
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 385)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Mammalia, Euthelia; Primates; Catarrhini; Homnidae; Homo.

Chisoe, S., Dietrich, N., Dubuque, T., Pavello, A., Gibb, W., Hawkins, M., Hulman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Matidis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)

TITLE JOURNAL MEDLINE
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 907 Std Error: 0.00
Seq primer: m13 -40 forward
High quality sequence stop: 276.
location/Qualifiers
1. 385

FEATURES
source
/organism="Homo sapiens"
/db_xref="GDB:3794746"
/db_xref="taxon:9606"
/clone="IMAGE:245500"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', ACTGGAAGAATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 88 a 88 c 102 g 106 t 1 others
ORIGIN

Query Match 54.2%; Score 271.2; DB 14; Length 385;
Best Local Similarity 93.0%; Pred. No. 5.9e-78;
Matches 306; Conservative 0; Mismatches 19; Indels 4; Gaps 2;

QY 66 ACTGAGAGACGAGACTGCTGATTTCTTAGCGCAGCTAAGATCCCTAGCCTGCTG 125
DB 327 ACTGAGAGACGAGACTGCTGATTTCTTAGCGCAGCTAAGATCCCTAGCCTGCTG 268

QY 126 GAAGGTGACGACGTCACCTTTAAACAGGGGGCTTGCACTAGCTCACCTGACCAAT 185
DB 267 GAAGGTGACGACGTCACCTTTAAACAGGGGGCTTGCACTAGCTCACCTGACCAAT 208

QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCAT 245
DB 207 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCAT 148

QY 246 CTATTGCTTGAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 305
DB 147 CTATTGCTTGAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 88

QY 306 CTGCAACAGAGAGGCCCCCTTTGGTGCTCCCTTTGATAGGAG--CTGTTTTCATGC 363
DB 87 C-GGCAACGGCAACCCCTTTGGTGCTCCCTTTGATAGGAGCTGTGTTTTCATGC 30

QY 364 TATTCACTCTATTAAATCTTGCA 392
DB 29 TATTCACTCTATTAAATCTTGCA 1

RESULT 12 342 bp mRNA linear EST 01-FEB-1995
T47345/c
LOCUS y71002.81 Stragagene placenta (#937225) Homo sapiens cDNA clone
DEFINITION IMAGE:70803 3', mRNA sequence.

ACCESSION T47345
VERSION T47345.1 GI:649326
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 342)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 10:07:23 ; Search time 1703.47 Seconds

(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501

Sequence: 1 gtcgcggccaccctccca.....ctgcatcatgcagggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: GenEmbl:.*
2: gb_da:.*
3: gb_htg:.*
4: gb_in:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_sts:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vi:.*
15: em_da:.*
16: em_fa:.*
17: em_hum:.*
18: em_in:.*
19: em_mu:.*
20: em_om:.*
21: em_or:.*
22: em_ov:.*
23: em_pat:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_sts:.*
28: em_un:.*
29: em_vi:.*
30: em_htg_hum:.*
31: em_htg_inv:.*
32: em_htg_other:.*
33: em_htg_mus:.*
34: em_htg_pln:.*
35: em_htg_rtd:.*
36: em_htg_mam:.*
37: em_htg_vit:.*
38: em_sy:.*
39: em_htgo_hum:.*
40: em_htgo_mus:.*
41: em_htgo_other:.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	10499	6 AX007980	AX007980 Sequence
2	498	99.4	56093	6 AX329572	AX329572 Sequence
3	498	99.4	56093	9 HSAC000064	AC000064 Human BAC
4	498	99.4	149194	9 AC007566	AC007566 Homo sapi
5	450	89.8	711	6 AX007998	AX007998 Sequence
6	434	86.6	174019	9 AP001538	AP001538 Homo sapi
7	434	86.6	340000	9 AP001674	AP001674 Homo sapi
8	431.2	86.1	163803	9 AC093531	AC093531 Homo sapi
9	426	85.0	839	9 AF127226	AF127226 Homo sapi
10	426	85.0	1329	6 AX001030	AX001030 Sequence
11	425.6	83.0	251124	9 HUA000660	AX000660 Homo sapi
12	415.6	83.0	251124	2 AC092510	AC092510 Papio cyn
13	409.6	81.8	187321	2 AC092510	AC092510 Homo sapi
14	409	81.6	83412	9 AC092843	AC092843 Homo sapi
15	407.2	81.3	140756	9 AL139038	AL139038 Human DNA
16	407.2	81.3	183499	2 AL607153	AL607153 Homo sapi
17	400.6	80.0	169029	2 CNS06C7R	AL359080 Human chr
18	400.2	79.9	110000	2 AL353584_0	AL353584 Homo sapi
19	400.2	79.9	170746	2 AL392173_0	AL392173 Human DNA
20	398.8	79.6	191863	9 AC010888	AC010888 Homo sapi
21	396.6	79.2	88328	9 AL357874	AL357874 Human DNA
22	396.4	79.1	157711	9 AC055738	AC055738 Homo sapi
23	396.4	79.1	207945	2 CNS01D5S	AL121841 Human chr
24	396.4	79.1	318442	2 AC021652	AC021652 Homo sapi
25	396.2	79.1	711	6 AX007997	AX007997 Sequence
26	392.8	78.4	114455	9 AC093742	AC093742 Homo sapi
27	391.4	78.1	165059	9 AC106856	AC106856 Homo sapi
28	390	77.8	137492	2 AL161721	AL161721 Homo sapi
29	389.8	77.8	181753	9 AC124075	AC124075 Homo sapi
30	389.6	77.8	143590	9 HS836J3	AL035706 Human DNA
31	388	77.4	119406	2 AC103595	AC103595 Homo sapi
32	387.4	77.3	76169	9 AC003014	AC003014 Human PAC
33	386.8	77.2	180635	9 AL360169	AL360169 Human DNA
34	386	77.0	145785	2 AC016173	AC016173 Homo sapi
35	385.8	77.0	161049	2 AL591842	AL591842 Homo sapi
36	385.4	76.9	111079	9 HS419C19	AL035407 Human DNA
37	385.2	76.9	156177	9 AC093896	AC093896 Homo sapi
38	384.8	76.8	42216	9 AC092899	AC092899 Homo sapi
39	384.8	76.8	170414	2 AC026019	AC026019 Homo sapi
40	384.6	76.8	119481	9 AL136234	AL136234 Human DNA
41	384.4	76.7	91097	9 AP001600	AP001600 Homo sapi
42	384.4	76.7	118241	9 AP001599	AP001599 Homo sapi
43	384.4	76.7	185018	9 AC110614	AC110614 Homo sapi
44	384.4	76.7	340000	9 AP001697	AP001697 Homo sapi
45	383	76.4	181850	9 AC018680	AC018680 Homo sapi

ALIGNMENTS

RESULT 1
AX007980
LOCUS AX007980 10499 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 3 from Patent WO9967395.
ACCESSION AX007980
VERSION AX007980.1 GI:9995677
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 9967395-A 3 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALTIEU, PATRICK M (FR)
FEATURES Location/Qualifiers

Source

1. 10499

BASE COUNT	3048 a	2676 c	2280 g	11495 t
ORIGIN				

Query Match	100.0%;	Score 5(1);	DB 6;	Length 10499;
Best Local Similarity	100.0%;	Pred. Nc. 6.2e-151;		
Matches 501; Conservative	0;	Mismatches	0;	Indels 0;

Qy	1	GTCTGGGCAACCTCCCAACAGACCTTAGGTTTTCCTGTGAGATGGGGACCTAGAGA	60
Db	9500	GTCTGGGCAACCTCCCAACAGACCTTAGGTTTTCCTGTGAGATGGGGACCTAGAGA	95558
Qy	61	CAGACTAGCTGGATTTCTTAGGCTGACTAAGA/TCCTTAGGCTTAGCTGGAGAGTAC	120
Db	9560	CAGACTAGCTGGATTTCTTAGGCTGACTAAGA/TCCTTAGGCTTAGCTGGAGAGTAC	96198
Qy	121	CACATCCACCTTTTAAACAGGGGGCTTGCAACTTAGCTCACCTGACCAATCAGAGACT	180
Db	9680	CACATCCACCTTTTAAACAGGGGGCTTGCAACTTAGCTCACCTGACCAATCAGAGACT	967978
Qy	181	CACATAAATGCTAATTAGGCAAGAACAAGAGGTAAAGAAATAGCCAAATCATATATGCTT	240
Db	9680	CACATAAATGCTAATTAGGCAAGAACAAGAGGTAAAGAAATAGCCAAATCATATATGCTT	97398
Qy	241	GAGAGCAACAGAGGAGGGACAATGATCGGGATAT/TAACCAAGTCMTGAGCCGGCAACG	300
Db	9740	GAGAGCAACAGAGGAGGGACAATGATCGGGATATTAACCAAGTCMTGAGCCGGCAACG	97998
Qy	301	GCAACCCCTTTTGGGTCCCTCCCTTTGTATGAGGAGCTCTGTATTTCATGCTATTTCACTC	360
Db	9800	GCAACCCCTTTTGGGTCCCTCCCTTTGTATGAGGAGCTCTGTATTTCATGCTATTTCACTC	98598
Qy	361	TATTTAAATCTTGGAACTGGCACTCTTCGGTGCATG/TTCTTAGCGGTTGAAGCTAGAGCTT	420
Db	9860	TATTTAAATCTTGGAACTGGCACTCTTCGGTGCATG/TTCTTAGCGGTTGAAGCTAGAGCTT	99198
Qy	421	CGCTCGGCATCACAACCTGCTGTTTGGCCGCAACCGAGACCGCGCTGACTCCATCC	480
Db	9920	CGCTCGGCATCACAACCTGCTGTTTGGCCGCAACCGAGACCGCGCTGACTCCATCC	99798
Qy	481	TCTGGATCATGCAAGGATGTCC	501
Db	9980	TCTGGATCATGCAAGGATGTCC	10000

RESULT 2

AX329572

LOCUS	AX329572	56093 bp
DEFINITION	Sequence 81 from Patent WO0194629.	

AX329572

VERS ION AX329572.1 GI:18102550

KEYWORDS

SOURCE
OPCGATEM

ORGANISM	CRANIATA	VERTEBRATA	EUTELEOSTOMI
<i>Homio sapiens</i>			
Eukaryota; Metazoa;			
Mammalia; Eutheria;			
Primates;			
Catarrhini; Homi-			
nidae; Homo.			

REFERENCE	AUTHORS
1	Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.

Horrigan, S., Soppet, D.R. and Weaver, Z.

TITLE Cancer gene determination and therapeutic screening using signature

gene bets
Patent: WO 0194629-A 81 13-DEC-2001;
JOURNAL

Avalon Pharmaceuticals (US)
Location/Qualifiers
FEATURES

BASE COUNT	16164	a	12346	c	10702	g	16881	t
ORIGIN								

Query Match	99.4%;	Score 498;	DB 6;	Length 56093;
Best Local Similarity	100.0%;	Pred. No. 7.2e-150;		
Matches 498; Conservative	0;	Mismatches	0;	Gaps 0;

QY	4	TCGGCCAACTCCCCAAGCAGCACTTAAGTTTTTCCTGTGAGATGGGGGACTAGAGACAG	63
Db	37504	TCGGCCAACTCCCCAAGCAGCACTTAAGTTTTTCCTGTGAGATGGGGGACTAGAGACAG	37563
QY	64	GACTAGCTGAGATTTCTTAGCTGACTTAAGATCCTTAAGCTTAGCTGGGAGGTGACAC	123
Db	37564	GACTAGCTGAGATTTCTTAGCTGACTTAAGATCCTTAAGCTTAGCTGGGAGGTGACAC	37623
QY	124	ATCCACCTTTTAAACACGGGGGCTTGCACTTAGCTCACAAGCTGACCAATCAGAGACTCAC	183
Db	37624	ATCCACCTTTTAAACACGGGGGCTTGCACTTAGCTCACAAGCTGACCAATCAGAGACTCAC	37683
QY	184	TAAAAATGCTAATTAGGCAAAAGCAGAGAGTAAAGAAATAGCCATCATCTATTGCTGAG	243
Db	37684	TAAAAATGCTAATTAGGCAAAAGCAGAGAGTAAAGAAATAGCCATCATCTATTGCTGAG	37743
QY	244	AGCAGAGCAGAGGGGACATGATGGGGATATTAACCCAAAGTTTGAGCCGGCAACGGCA	303
Db	37744	AGCAGAGCAGAGGGGACATGATGGGGATATTAACCCAAAGTTTGAGCCGGCAACGGCA	37803
QY	304	ACCCCTTTTGAGGTCCCTTCCTCTTTGATGGAGCTCTGTGTTTCATGCTATTTCACTTAT	363
Db	37804	ACCCCTTTTGAGGTCCCTTCCTCTTTGATGGAGCTCTGTGTTTCATGCTATTTCACTTAT	37863
QY	364	TAAATCTTGCACTGCACTCTTCTGTGTCAGATTTCTTAACGGCTTGAGCTGAGCTTGGC	423
Db	37864	TAAATCTTGCACTGCACTCTTCTGTGTCAGATTTCTTAACGGCTTGAGCTGAGCTTGGC	37923
QY	424	TCGCCATCCACCACTGCTGTTTGGCCGCAACCGCAGACCCGCGCTGACTCCCATCCCTT	483
Db	37924	TCGCCATCCACCACTGCTGTTTGGCCGCAACCGCAGACCCGCGCTGACTCCCATCCCTT	37983
QY	484	GGATCATGACGGGTGCC	501
Db	37984	GGATCATGACGGGTGCC	38001

RESULT 3

HSAC000064

LOCUS	HSAC000064	56093 bp	DNA	Linear	PRI 13-
DEFINITION	Human BAC clone RG083M05 from 7q21-7q22, complete sequence				

ACCESSION	AC000064
IMPACTOR	15000004

```

VERSION          AC0000064.1  GL:1669362
VERWORD          YMD

```

KEYWORDS

SOURCE

ORIGINALS

Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Hominidae: Homo.

REFERENCE
1 (bases 1 to 56093)

ALITHORS Paul Lev. A.

TITLE	SEQUENCE OF H. sapiens BAC clone RG083M05
	The sequence of H. sapiens BAC clone RG083M05

JOURNAL Unpublished (1996)

REFERENCE 2 (bases 1 to 56093)

AUTHORS Waterston, R.

TITLE Direct Submission

JOURNAL Submitted (13-NOV-1995)

COMMENT **Genome Sequencing Cer**

Department of Gene

St. Louis, MO 6310

e-mail: sapiens@wa

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The
library contains cloned DNA from a human male fibroblast cell line
978SK. For references see: Shitaya et al., Proc. Natl. Acad. Sci.
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBelo
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this
clone is at base position 1 of H_RG083M05; actual end is at 56093
of H_RG083M05

This clone contains STS SMS1725.

FEATURES

source

Location/Qualifiers

1..56093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"

/map="7q21-7q22"
/clone_id="H_RG083M05"
/clone_1id="CTTB-978SK-B"
complement(838..1131)
/rpt_family="ALU"

repeat_region

gene

CDS

/gene="WUGSC:H_RG083M05.1"
join(<1360..1503,4181..4370,4587..4774,6422..6556,
9483..9547,11631..11773,11864..12021,13131..13296,
14885..14988,16349..16546,16837..16971)
/gene="WUGSC:H_RG083M05.1"
/note="Atpase; strong similarity to peroxisome
biosynthesis protein PAB1 (PID:g1172019); coded for by
human cDNA C04279 (NID:g1467530)"

/codon_start=1
/protein_id="AAB46346.1"
/db_xref="GI:1669371"

/translation="KRLENIQKLEVAFAEAWMPQPSVLLDLDLJAGPAVEHEH
SPAEORCEILCNVTKKDKDINKFTDLOHAKETGFGVADPFLVDRAHSL
SRGSIQREKLVLTLDPOKALRGFLPASLRSVNHKPRDLMGKLGHEVQVIMD
TIQPAVCLKEKTEPFLFANLPTRGTGILYGPPTGKTLAIVTARSRNFTISV
KGPGLSKYIGASQAVADIFIRQAAPCLLFEDEFESIAPRGHNTGTVREVNQ
LITQIDVEGIGQGVYLAATSRPDLIPALRGLRDLKCYCPCPDVITISYLSKTO
QMHSFVSRLEINLVLSPLADVDLQHVAVTDSFGADKALKLYNAOLALHG
MLSKMSEILLPDESKFMWRLYFGSSYSELENGTSDLSQCLASPSMTODLGPV
CKQQLFQPPYLRASQEGCELTORORDLRADISIIKRYRSGSEDESMNPGFI
KTRLAISQSHMTALGHTRPSISIEDWKNFAEL"

repeat_region

repeat_region

repeat_region

repeat_region

misc_feature

misc_feature

misc_feature

/note="match to human 3' EST H75782 (NID:g1049794), bases
287-444"
/rpt_family="ALU"
complement(4948..5130)
/rpt_family="ALU"
complement(6581..7133)
/rpt_family="L1"
complement(17767..8037)
/rpt_family="ALU"
complement(8186..8472)
/rpt_family="ALU"
8473..8625
/gene="WUGSC:H_RG083M05.1"
/note="match to human 3' EST H75782 (NID:g1049794), bases
287-444"
8841..9161
/gene="WUGSC:H_RG083M05.1"
/note="match to human 5' EST H75921 (NID:g1050050), bases
21-348"
9481..9547
/gene="WUGSC:H_RG083M05.1"
/note="match to human 5' EST N22627 (NID:g1130501), bases
276-343"

repeat_region

misc_feature

repeat_region

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

misc_feature

repeat_region

misc_feature

exon

repeat_region

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

complement(12612..12907)
/rpt_family="ALU"
13670..13793
/gene="WUGSC:H_RG083M05.1"
/note="match to human 5' EST H41382 (NID:917434), bases
143-266"

13794..13877
/rpt_family="ALU"
13878..13906
/gene="WUGSC:H_RG083M05.1"
/note="match to human 5' EST H41382 (NID:917434), bases
30-58"

13907..14104
/rpt_family="ALU"
complement(14110..14137)
/rpt_family="L1"
complement(15618..15907)
/rpt_family="ALU"
17227..17522
/rpt_family="ALU"
18667..19235
/note="match to human fetal brain 5' EST D61494
(NID:970409), bases 1-255, and to human 3' EST R07476
(NID:9759399)"

19550..19670
/rpt_family="ALU"
21507..37303
/note="similarity to various SS-RNA virus polyproteins;
pseudogene; region of matches and close matches to
multiple human ESTs, see R68740 (NID:9842257)"
37316..37489
/note="Grail prediction, score = 80"

/evidence=not experimental
complement(38538..39224)
/rpt_family="ALU"
39225..39707
/note="match to multiple human ESTs, see N30113
(NID:91148633)"
39800..40085
/rpt_family="ALU"

complement(40247..40538)
/rpt_family="ALU"
complement(40632..40924)
/rpt_family="ALU"
complement(42283..42891)
/rpt_family="ALU"
complement(45474..45613)
/rpt_family="ALU"
complement(45614..45737)
/note="match to human 3' EST H48698 (NID:9988738), bases
125-333"

complement(46107..47026)
/note="match to multiple human ESTs, see N61064
(NID:91243765), H48697 (NID:9988737), and W78831
(NID:9273146)"
complement(47027..47318)
/rpt_family="ALU"
complement(47365..47782)
/note="match to multiple human ESTs, see W37495
(NID:91319089)"
47898..48115
/note="match to human 5' EST H62306 (NID:g1015138), bases
93-368"

complement(48116..48405)
/rpt_family="ALU"
complement(48406..48584)
/note="match to human 3' EST N29952 (NID:g1148472), bases
290-455, and 5' EST R12730 (NID:9765806)"
complement(48787..49405)
/rpt_family="ALU"
complement(49406..49534)
/note="match to human 3' EST R65794 (NID:9838432), bases
309-440"

```

repeat_region      complement(49638..49672)
                    /rpt family="ALU"
misc_feature        complement(49674..49890)
                    /note="match to human 3' EST N29952 (NID:g1148472) and 5'
                    EST N29938 (NID:g1148438), sequences are from opposite
                    ends of the same clone"
gene                complement(49698..51806)
                    /gene="WUSC:H.RG083M05.2"
                    complement(join(49698..49888,51575..51806))
                    /note="WUSC:H.RG083M05.2"
                    /note="coded for by human cDNA #37388 (NID:g1319205),
                    R6591 (NID:g838529), R65794 (NID:g838432) and R65794
                    (NID:g838432)"
CDS                /codon_start=1
                    /protein_id="AAB46345.1"
                    /db_xref="GI:1669370"
                    /translation="MFYFPGGILFPCPGVYVQIGVSVDEQDKRYAARGR
                    IDDYCEKSAALWLIPTLSPPRQDFDASTITGEEDLPKMEYLEFVCHAPSEYFK
                    SRSSPFPVPTPRKGYIMTVHGJTPALTIKESVANHL"
exon                complement(51576..51751)
                    /gene="WUSC:H.RG083M05.2"
                    /note="Grail prediction, score = 86"
repeat_region      /evidence="not experimental"
                    complement(52052..52325)
                    /rpt family="I1"
                    /note="match to human EST M79192 (NID:g273505) base 2-289"
misc_feature        55557..55843
Query Match        99.4%; Score 498; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. No. 7,2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

ACCESSION          AC007566
VERSION            AC007566.2
KEYWORDS           GI:11181861
SOURCE             HTG.
ORGANISM           Homo sapiens.
                   Homo sapiens.
REFERENCE           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS            1 (bases 1 to 149194)
                   Sulston, J.B. and Waterston, R.
TITLE              Toward a complete human genome sequence
JOURNAL            Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE            99063792
PUBMED             9847074
REFERENCE           Du, Z.
                   2 (bases 1 to 149194)
TITLE              The sequence of Homo sapiens BAC clone CTB-10G5
JOURNAL            Unpublished (2001)
REFERENCE           3 (bases 1 to 149194)
                   Waterston, R.H.
AUTHORS            Direct Submission
JOURNAL            Submitted (15-MAY-1999) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE           4 (bases 1 to 149194)
                   Waterston, R.
AUTHORS            Direct Submission
JOURNAL            Submitted (02-OCT-2000) Department of Genetics, Washington
                   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE           5 (bases 1 to 149194)
                   Waterston, R.H.
AUTHORS            Direct Submission
JOURNAL            Submitted (16-NOV-2000) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE           6 (bases 1 to 149194)
                   Waterston, R.H.
AUTHORS            Direct Submission
JOURNAL            Submitted (03-JAN-2002) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE           7 (bases 1 to 149194)
                   Waterston, R.H.
AUTHORS            Direct Submission
JOURNAL            Submitted (06-FEB-2002) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE           8 (bases 1 to 149194)
                   Waterston, R.
AUTHORS            Direct Submission
JOURNAL            Submitted (01-MAR-2002) Department of Genetics, Washington
                   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
                   On Nov 16, 2000 this sequence version replaced gi:4835815.
COMMENT            ----- Genome Center
                   Center: Washington University Genome Sequencing Center
                   Center code: WUGSC
                   Web site: http://genome.wustl.edu/gsc
                   Contact: saplens@watson.wustl.edu
                   ----- Summary Statistics
                   -----
                   Center project name: H_RG010050

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.regen.com>).
VECTOR: pBelOAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES
source 1..149194
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_1lb="CTB-978SK-B"
1..1634
/ftc_family="L1"
repeat_region 2248..2388
/note="match to EST BG752883 (NID:g14063536)"
misc_feature 2248..2387
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 2248..2387
/note="match to EST AM579261 (NID:g7254310)"
misc_feature 2248..2387
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 2248..2387
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 2248..2374
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
2248..2287
/note="match to EST AV686676 (NID:g10288539)"
misc_feature 2253..2387
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 2344..2387
/note="match to EST AV686676 (NID:g10288539)"
repeat_region 2696..3066
/ftc_family="MALR"
repeat_region 3108..3392
/ftc_family="Alu"
misc_feature 3540..3628
/note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
misc_feature 3540..3628
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature 3542..3632
/note="match to EST AV686676 (NID:g10288539)"
misc_feature 3542..3628
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 3542..3628
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 3542..3628
/note="match to EST BG766882 (NID:g14077535)"

misc_feature 3542..3628
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 3542..3609
/note="match to EST AM579261 (NID:g7254310)"
misc_feature 3562..3628
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 3598..3628
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3714..3785
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 3717..3785
/note="similar to Mus musculus EST BE994936 (NID:g10678674)"
misc_feature 3719..3785
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 3719..3785
/note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
misc_feature 3719..3785
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 3719..3785
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 3719..3785
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3719..3785
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 3719..3785
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 3719..3785
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature 3860..3862
/note="match to EST AV686676 (NID:g10288539)"
repeat_region 4730..4827
/ftc_family="L2"
repeat_region 4834..4925
/ftc_family="L1"
misc_feature 5715..5860
/note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
misc_feature 5715..5860
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 5715..5860
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 5715..5860
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 5715..5860
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 5715..5860
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 5715..5860
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 5715..5860
/note="similar to Mus musculus EST BE994936

Query Match 99.4%; Score 498; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 8.2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCCAAGCTCCCAACAGACAGCTAGCTTTCTGTTGAGATGGGGAGCTGAGAGCAG 63
DB 84500 TCGGCCAAGCTCCCAACAGACAGCTAGCTTTCTGTTGAGATGGGGAGCTGAGAGCAG 84441
QY 64 GACTAGCTGATTTCTAGAGCTGACTAAGATCCCTAAGCTGCTGGAAGTGACAC 123
DB 84440 GACTAGCTGATTTCTAGAGCTGACTAAGATCCCTAAGCTGCTGGAAGTGACAC 84381
QY 124 ATCCACCTTTAAACAGCGGGCTTGCACTTAGCTCAACCTGACCAATCAGAGAGCTCAC 183
DB 84380 ATCCACCTTTAAACAGCGGGCTTGCACTTAGCTCAACCTGACCAATCAGAGAGCTCAC 84321
QY 184 TAAATGCTAATTAAGGCAAGACAGAGGTAAAGAAATACCAATCATCTATTGCTGAG 243
DB 84320 TAAATGCTAATTAAGGCAAGACAGAGGTAAAGAAATACCAATCATCTATTGCTGAG 84261

```
QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATATAA(CCAAGTCTTGAGCCGGCAACGGCA 303
DB 84260 AGCAGAGAGAGGAGCAATGATCGGATATATAA(CCAAGTCTTGAGCCGGCAACGGCA 84201
QY 304 ACCCCCTTTGGGTCCCTCCCTTTGATAGGAGCCTGTGTTTTCATGCTATTTCTCTAT 363
DB 84200 ACCCCCTTTGGGTCCCTCCCTTTGATAGGAGCCTGTGTTTTCATGCTATTTCTCTAT 84141
QY 364 TAAATCTGCAACTGACTCTTCTGATGCTATGTTTCTTACGCTTTGAGCTGAGCTTTGCG 423
DB 84140 TAAATCTGCAACTGACTCTTCTGATGCTATGTTTCTTACGCTTTGAGCTGAGCTTTGCG 84081
QY 424 TCGGCATCCACCACTGCTGTTTCCGCGCAGCCGAGACCCGCGCTGACTCCCATCCCTCT 483
DB 84080 TCGGCATCCACCACTGCTGTTTCCGCGCAGCCGAGACCCGCGCTGACTCCCATCCCTCT 84021
QY 484 GGATCATGACGAGGTGTCC 501
DB 84020 GGATCATGACGAGGTGTCC 84003

RESULT 5
AX007998 711 bp DNA linear PAT 06-SEP-2000
LOCUS AX007998
DEFINITION Sequence 21 from Patent WO9967395.
ACCESSION AX007998
VERSION AX007998.1 GI:995695
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 711)
Perrin, J.P., Rieger, P. and Alliel, P.M.
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
Patent: WO 9967395-A 21 29-DEC-1995;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES
SOURCE
1. 711
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 171 a 209 c 160 g 171 :
ORIGIN
Query Match 89.8%; Score 450; DB 6; Length 711;
Best Local Similarity 100.0%; Pred. No. 1,0e-134; Indels 0; Gaps 0;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 ACTGAGAGCAGAGCTAGCTGATTTCTTGAAGCTCCCTAAGCTGAGCTGAG 111
DB 1 ACTGAGAGCAGAGCTAGCTGATTTCTTGAAGCTCCCTAAGCTGAGCTGAGCTGAG 60
QY 112 GAAGGTGACCACTCCATTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAAT 171
DB 61 GAAGGTGACCACTCCATTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAAT 120
QY 172 CAGAGAGCTCAATAATCTTAATTTAGGCAAGACAGGAGGTAAGAAATAGCAATCT 231
DB 121 CAGAGAGCTCAATAATCTTAATTTAGGCAAGACAGGAGGTAAGAAATAGCAATCT 180
QY 232 CTATGCTGAGAGCAGAGCAGAGGAGCAATGATGGAATTAATTAACCAAGCTTTCAG 291
DB 181 CTATGCTGAGAGCAGAGCAGAGGAGCAATGATGGAATTAATTAACCAAGCTTTCAG 240
QY 292 CCGGCAACCGGCAACCCCTTTGGGTCCCTCCCTTTGT, TGGAGGCTCTGTTTTCATGCT 351
DB 241 CCGGCAACCGGCAACCCCTTTGGGTCCCTCCCTTTGT, TGGAGGCTCTGTTTTCATGCT 300
QY 352 ATTTGACTTATTAATCTTTGCACTGCACTGCTTCTGCTCATGTTTCTTAAGGCTTGA 411
DB 301 ATTTGACTTATTAATCTTTGCACTGCACTGCTTCTGCTCATGTTTCTTAAGGCTTGA 360
```

```
QY 412 CTGAGCTTTGCTGCGCATCCACCACTGCTGTTTGGCGCCACCGGACAGACCCGCCGCTGAC 471
DB 361 CTGAGCTTTGCTGCTGCGCATCCACCACTGCTGTTTGGCGCCACCGGACAGACCCGCCGCTGAC 420
QY 472 TCCCATCCCTCTGGATCATGAGGAGGTCC 501
DB 421 TCCCATCCCTCTGGATCATGAGGAGGTCC 450

RESULT 6
AP001538/c 174019 bp DNA linear PRI 25-MAR-2000
LOCUS AP001538
DEFINITION Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B853K11,
IL56-APP region, complete sequence.
ACCESSION AP001538
VERSION AP001538.1 GI:7328982
KEYWORDS
SOURCE HTG.
ORGANISM Homo sapiens DNA, clone:B853K11.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 174019)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Torok, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 174,019 genomic DNA of 21q21.1-q21.2
Published Only in Database (2000)
2 (bases 1 to 174019)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Torok, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (23-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kilbasato Univ., 1-15-1 Kilbasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gs.c.riken.go.jp, Tel:81-42-778-9923,
URL:http://hgp.gs.c.riken.go.jp/, Fax:81-42-778-9924)
FEATURES
SOURCE
1. 174019
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1-q21.2"
/clone="B853K11"
BASE COUNT 53303 a 31416 c 32248 g 57052 t
ORIGIN
Query Match 86.6%; Score 434; DB 9; Length 174019;
Best Local Similarity 92.0%; Pred. No. 4.9e-129; Indels 0; Gaps 0;
Matches 458; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGACTGAGAGACAG 63
DB 36468 TTGGCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGACTGAGAGACAG 36409
QY 64 GACTAGCTGATTTCTGAGCTGACTAAGAAATCCCTAAGCTAGCTGGAAAGGTACAC 123
DB 36408 GACTAGCTGATTTCTGAGCTGACTAAGAAATCCCTAAGCTAGCTGGAAAGGTACAC 36349
QY 124 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGACTCAC 183
DB 36348 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGACTCAC 36289
QY 184 TAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG 243
DB 36288 TAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG 36229
QY 244 AGCAGCAGAGAGGAGCAATGATGGAATTAATTAACCAAGCTTTCAGAGCCGCAACGGCA 303
DB 36228 AGCAGCAGAGAGGAGCAATGATGGAATTAATTAACCAAGCTTTCAGAGCCGCAACGGCA 36169
QY 304 ACCCCCTTTGGGTCCCTCCCTTTGATAGGAGCTCTGTTTTCATGCTATTTCTCTAT 363
```



```

/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R28M9, 3' partial"
/clone_id="RRC1-11 BAC library"
/note="Accession No. AP000431"
311751..3540000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="c103E2331, 3' partial"
/clone_id="ICRFc103 Coe mid library"
/note="Accession No. AL132933"
15..444
/note="L1MA4A"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
555..3531
/note="L1MA4A"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
3770..3876
/note="L2"
/rpc_family="LINE/L2"
/rpc_type=DISPERSED
5645..6178
/note="L1MA2"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
6181..6534
/note="THE1B"
/rpc_family="LTR/MaLR"
/rpc_type=DISPERSED
8021..8370
/rpc_family="L2"
/rpc_type=DISPERSED
10213..10520
/note="L1MA2"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
15810..15851
/note="(CA)n"
/rpc_family="Simple_repeat"
/rpc_type=TANDEM
complement(16628..16697)
/note="MER53"
/rpc_family="DNA"
/rpc_type=DISPERSED
16638..16715
/note="MER53"
/rpc_family="DNA"
/rpc_type=DISPERSED
18847..19585
/note="L2"
/rpc_family="LINE/L2"
/rpc_type=DISPERSED
19795..19917
/note="(TA)n"
/rpc_family="Simple_repeat"
/rpc_type=TANDEM
20023..20166
/note="(TA)n"
/rpc_family="Simple_repeat"
/rpc_type=TANDEM
20285..20448

```

Query Match 86.6%; Score 434; DB 9; Length 340000;

Best Local Similarity 92.0%; Pred. No. 5,4e-129; Matches 458; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

QY 4 TCGGCAACCTCCCAACAGACCTTAGGTTTCTGTTAGATGGGAGCTAGAGACAG 63
Db 27679 TTGGCAACCTCCCAACAGACCTTAGGTTTCTGTTAGATGGGAGCTAGAGACAG 27620
QY 64 GACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTAAGCTAGCTGGAGAGTACAC 123
Db 27619 GACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTAAGCTAGCTGGAGAGTACAC 27560
QY 124 ATCCACTTTTAAACGGGGCTTGAATTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 183
Db 27559 ATCCACTTTTAAACGGGGCTTGAATTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 27500
QY 184 TAAATGCTAATTAGGCAAGACAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 243
Db 27439 TAAATGCTAATTAGGCAAGACAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 27440
QY 244 AGCAGCAGAGAGGAGCAATGATGCGATATMAACCAAGTCTTGCAGCCGCAACGCA 303
Db 27439 AGCAGCAGAGAGGAGCAATGATGCGATATMAACCAAGTCTTGCAGCCGCAACGCA 27380
QY 304 ACCCCCTTTGGGTCCCTCCCTTTGTATAGGAGCTCTGTTTCAATGCTATTTACTAT 363
Db 27379 ACCCCCTTTGGGTCCCTCCCTTTGTATAGGAGCTCTGTTTCAATGCTATTTACTAT 27320
QY 364 TAAATCTGCAACTGCACTCTTCTGCTCCATGTTTCTTAAGGCTTGAAGCTGCTTGC 423
Db 27319 TAAATCTGCAACTGCACTCTTCTGCTCCATGTTTCTTAAGGCTTGAAGCTGCTTGC 27260
QY 424 TCGCATTCACCACTGCTGTTTCCGCGCACCGACAGCCCGCGTACTCCATCTCT 483
Db 27259 TCGCATTCACCACTGCTGTTTCCGCGCGTGTGCGAGACCTGCGACACTGCCATCTCT 27200
QY 484 GGATCATGCAAGGTGCTCC 501
Db 27199 GGATCATGCAAGGTGCTCC 27182

```

RESULT 8
AC093531/c 163803 bp DNA linear PRI 16-NOV-2001
LOCUS Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
DEFINITION AC093531
ACCESSION AC093531.2 GI:16945981
VERSION
KEYWORDS HUG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 163803)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 163803)
REFERENCE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 163803)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
AUTHORS Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 16, 2001 this sequence version replaced gi:15383820.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.
Location/Qualifiers

source 1. .163803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-405L7"

BASE COUNT 54797 a 31372 c 29086 g 48548 t
ORIGIN

Query Match 86.1%; Score 431.2; DB 9; Length 163803;
Best Local Similarity 91.4%; Pred. No. 4e-128;
Matches 457; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 2 TCTGGCCAACTCCCAAGACAGCACTTAAGTTTCTGTTGAGTGGGGAGCTAGAGAC 61
Db 117108 TGTACGCAACTCCCAAGACAGCACTTAAGTTTCTGTTGAGTGGGGAGCTAGAGAC 117049

QY 62 AGAGCTAGCTGATTTCTTGGCTGAGTAAATCCCTAAGCTAGCTGGGAAGTAC 121
Db 117048 AGAGCTAGCTGATTTCTTGGCTGAGTAAATCCCTAAGCTAGCTGGGAAGTAC 116989

QY 122 ACATCACTTTAAACAAGGGGCTTGCACCTTACACCTGACCAATCAGAGAGCTC 181
Db 116988 GCATCCATCTTTAAACAAGGGGCTTGCACCTTACACCTGACCAATCAGAGAGCTC 116929

QY 182 ACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCCATCATCTATTGCTG 241
Db 116928 ACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCCATCATCTATTGCTG 116869

QY 242 AGAGCAAGAGAGAGGAGCAATGATCGGATATAACCCAGTCTTGGAGCGGCAACGG 301
Db 116868 AGAGCAAGAGAGAGGAGCAATGATCGGATATAACCCAGTCTTGGAGCGGCAACAG 116809

QY 302 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTCATGCTATTCACTCT 361
Db 116808 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTCATGCTATTCACTCT 116749

QY 362 ATTAATCTTGCAACTGCACTCTTCTGCTCAGTCTTCTTAAGGCTTGAAGTGA 421
Db 116748 ATTAATCTTGCAACTGCACTCTTCTGCTCAGTCTTCTTAAAGTGAAGTGA 116689

QY 422 GCTGGCATCAGACACGCTGCTTTGGCCGACCGGAGACCGCGCTGACTCCATCCCT 481
Db 116688 GCTGGCATCAGACACGCTGCTTTGGCCGACCGGAGACCGCGCTGACTCCATCCCT 116629

QY 482 CTGGATCATGAGGGTGCTCC 501
Db 116628 TTGGATCCAGCAGAGTGCTCC 116609

RESULT 9
AF127226 839 bp DNA linear PRI 17-JUN-1999
LOCUS AF127226
DEFINITION Homo sapiens Human endogenous retrovirus W, 3' long terminal
repeat, partial sequence.
ACCESSION AF127226
VERSION AF127226.1 GI:5081468.
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Cumanian-Paraz, A., Komurian-Pradel, F., Ott, C., Rajorharrison, A. and
Perron, H.
TITLE MSRV-related HERV-W elements in human DNA
JOURNAL Unpublished
AUTHORS (bases 1 to 839)
Cumanian-Paraz, A., Komurian-Pradel, F., Ott, C., Rajorharrison, A. and
Perron, H.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-1999) UMR103 CNRS-Biomerieux, Biomerieux, 46,
Allee D' Italie, Lyon 69007, France
FEATURES
Location/Qualifiers

source 1. .839
/organism="Homo sapiens"
/db_xref="taxon:9606"
LRR <1. .>839
/note="3' long terminal repeat; U3, R and U4 regions"

repeat_region 1. .839
/note="MSRV-related"
/rpt_family="Human endogenous retrovirus W"
/rpt_type=dispersed

BASE COUNT 205 a 233 c 192 g 206 t 3 others
ORIGIN

Query Match 85.0%; Score 426; DB 9; Length 839;
Best Local Similarity 90.6%; Pred. No. 9.9e-127;
Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGCCAACTCCCAAGACAGCACTTAAGTTTCTGTTGAGTGGGGAGCTAGAGAC 61
Db 4 TGTACGCAACTCCCAAGACAGCACTTAAGTTTCTGTTGAGTGGGGAGCTAGAGAC 63

QY 62 AGAGCTAGCTGATTTCTTGGCTGAGTAAATCCCTAAGCTAGCTGGGAAGTAC 121
Db 64 AGAGCTAGCTGATTTCTTGGCTGAGTAAATCCCTAAGCTAGCTGGGAAGTAC 123

QY 122 ACATCACTTTAAACAAGGGGCTTGCACCTTACACCTGACCAATCAGAGAGCTC 181
Db 124 GCATCCATCTTTAAACAAGGGGCTTGCACCTTACACCTGACCAATCAGAGAGCTC 183

QY 182 ACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCCATCATCTATTGCTG 241
Db 184 ACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCCATCATCTATTGCTG 243

QY 242 AGAGCAAGAGAGAGGAGCAATGATCGGATATAACCCAGTCTTGGAGCGGCAACGG 301
Db 244 AGAGCAAGAGAGAGGAGCAATGATCGGATATAACCCAGTCTTGGAGCGGCAACAG 303

QY 302 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTCATGCTATTCACTCT 361
Db 304 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTCATGCTATTCACTCT 363

QY 362 ATTAATCTTGCAACTGCACTCTTCTGCTCAGTCTTCTTAAGGCTTGAAGTGA 421
Db 364 ATTAATCTTGCAACTGCACTCTTCTGCTCAGTCTTCTTAAAGTGAAGTGA 423

QY 422 GCTGGCATCAGACACGCTGCTTTGGCCGACCGGAGACCGCGCTGACTCCATCCCT 481
Db 424 GCTGGCATCAGACACGCTGCTTTGGCCGACCGGAGACCGCGCTGACTCCATCCCT 483

QY 482 CTGGATCATGAGGGTGCTCC 501
Db 484 TTGGATCCAGCAGAGTGCTCC 503

RESULT 10
AX001030 1329 bp DNA linear PAT 10-MAR-2000
LOCUS AX001030
DEFINITION Sequence 12 from Patent WO9902666.
ACCESSION AX001030
VERSION AX001030.1 GI:7241266
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1329)
AUTHORS Ott, C. and Bedin, F.
TITLE RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR
ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR
DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
JOURNAL PATENT: WO 9902666-A 12 21-JAN-1999;
BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES
Location/Qualifiers
1. .1329
/organism="unidentified"

BASE COUNT 346 a /db xref="taxon:32644"
 ORIGIN 352 c 288 g 340 t 3 others

Query Match 85.0%; Score 426; DB 6; Length 1329;
 Best Local Similarity 90.6%; Pred. No. 1=126;
 Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGGCAACCTCCCAAGACACTAGTTTCTGTGTGAGATGGGGGAGCTGAGAGAC 61
 DB 497 TGTGACCAACCTCCCAAGACTAGTTTCTGTGTGAGAGGGTGAAGAGAGAC 556

QY 62 AGAAGTCTGATTTCTAGGCTGACTAAGATCTTAAAGCTTGAAGTGGAGAGTAC 121
 DB 557 AGAAGTCTGATTTCTAGGCTGACTAAGATCTTAAAGCTTGAAGTGGAGAGTAC 616

QY 122 ACATTCACCTTTAAACAGAGGGGCTTGCAATGCTCAACCTGACCAATGAGAGCTC 181
 DB 617 GCATTCATCTTTAAACAGAGGGGCTTGCAATGCTCAACCTGACCAATGAGAGCTC 676

QY 182 ACTAAATGCTAATTAAGCAAGAGAGGTAAGAAATAGCCATGATCTATTTGCTG 241
 DB 677 ACTAAATGCTAATTAAGCAAGAGAGGTAAGAAATAGCCATGATCTATTTGCTG 736

QY 242 AGAGCAGCAGAGAGAGGCAATGATCGGATATTAATCCAACTTTCAGCCGCAACGG 301
 DB 737 AGAGCAGCAGAGAGAGGCAATGATCGGATATTAATCCAACTTTCAGCCGCAACGG 796

QY 302 CAACCCCTTTGGGTCCTCCCTCTTTGATAGGAGCCCTGTTTATGATGATTTACTCT 361
 DB 797 CAACCCCTTTGGGTCCTCCCTCTTTGATAGGAGCCCTGTTTATGATGATTTACTCT 856

QY 362 ATTAATCTTGCACTGCACTCTTCTGTCATGTTTCTTAAGGCTTGAAGCTGAGCTTTC 421
 DB 857 ATTAATCTTGCACTGCACTCTTCTGTCATGTTTCTTAAGGCTTGAAGCTGAGCTTTC 916

QY 422 GCTGCGCATCCACCATGCTGTTTCCCGCACGAGACCCGCTGATCTCCATCCCT 481
 DB 917 GCTGCGCATCCACCATGCTGTTTCCCGCACGAGACCCGCTGATCTCCATCCCT 976

QY 482 CTGATCATGAGGCTGTC 501
 DB 977 TTGATCCAGAGAGTGTCC 996

RESULT 11
 AX000970 783 bp 1MA linear PAT 10-MAR-2000
 LOCUS Sequence 15 from Patent WO9902696.
 ACCESSION AX000970
 VERSION AX000970.1 GI:7241212
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 783)
 AUTHORS Beseme, F. and Blond, J.
 TITLE ENDOGENETIC RETROVIRAL SEQUENCES ASSOCIATED WITH AUTOIMMUNE
 DISEASES OR WITH PREGNANCY DISORDERS
 JOURNAL Patent: WO 9902696-A 15 21-JAN-1999;
 BIO MEDICX (FR); BESEME FREDERIC (FR)
 FEATURES 1..783
 source Location/Qualifiers
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 173 a 213 c 166 g 180 t 51 others
 ORIGIN
 Query Match 85.0%; Score 425.6; DB 6; Length 783;
 Best Local Similarity 94.0%; Pred. No. 1.3e-126;
 Matches 422; Conservative 25; Mismatches 1; Indels 1; Gaps 1;

QY 54 TGAGAGACAGACTAGTGTATTTCTAGGCTGAGTAAATCCCTTAAGCTAGCTGGGA 113

DB 1 TGAGAGACAGACTAGTGTATTTCTAGGCTGAGTAAATCCCTTAAGCTAGCTGGGA 60

QY 114 AGGTGACCAATCCACTTTAAACACGGGGCTTGCAACTTAGCTACACCTGACCAATCA 173
 DB 61 AGGTGACCAATCCACTTTAAACACGGGGCTTGCAACTTAGCTACACCTGACCAATCA 120

QY 174 GAGAGCTCACTAAATGCTAATTTAGGCAAGAGAGGTGAAAGAAATATACCAATCACT 233
 DB 121 GAGAGCTCACTAAATGCTAATTTAGGCAAGAGAGGTGAAAGAAATATACCAATCACT 180

QY 234 ATTGCTGAGAGCAGACAGAGAGGCAATGATCGGATATTAACCAAGCTTTCAGACC 293
 DB 181 ATTGCTGAGAGCAGACAGAGAGGCAATGATCGGATATTAACCAAGCTTTCAGACC 240

QY 294 GGCACGACA-ACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTGTTTTCATCTA 352
 DB 241 GGCACGACAGMCCCTTTGGGTCCCTCCCTTTGATAGGAGCTGTTTTCATCTA 300

QY 353 TTTGACTGATTAATCTTGCAATGCACTCTTGTGTCATGTTTCTTAAGGCTTGAAC 412
 DB 301 TTTGACTGATTAATCTTGCAATGCACTCTTGTGTCATGTTTCTTAAGGCTTGAAC 360

QY 413 TGAGCTTTCGCTGGCATCCACCATGCTGTTTGGCGCACCGGACCCGCTGACT 472
 DB 361 TGAGCTTTCGCTGGCATCCACCATGCTGTTTGGCGCACCGGACCCGCTGACT 420

QY 473 CCATCCCTCTGATCATGAGGCTGTC 501
 DB 421 CCATCCCTCTGATCATGAGGCTGTC 449

RESULT 12
 HUAE000660 251124 bp DNA linear PRI 15-JUN-2001
 LOCUS Homo sapiens T-cell receptor alpha delta locus from bases 501613 to
 DEFINITION 752736 (section 3 of 5) of the Complete Nucleotide Sequence.
 ACCESSION AE000660 AE000521 U85197
 VERSION AE000660.1 GI:2358042
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 251124)
 AUTHORS Koop, B.F., Rowen, L., Wang, K., Kuo, C.-L., Seto, D., Lemstra, J.A.,
 Howard, S., Shan, W., Deshpande, P. and Hood, L.
 TITLE The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region:
 organization, sequence, and evolution of 97.6 kb of DNA
 JOURNAL Genomics 19 (3), 478-493 (1994)
 MEDLINE 94245236
 PUBMED 8188290
 REFERENCE 2 (bases 1 to 251124)
 AUTHORS Boyesen, C., Simon, M.I. and Hood, L.
 TITLE Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with
 bacterial artificial chromosome clones
 JOURNAL Genome Res. 7 (4), 330-338 (1997)
 MEDLINE 97264339
 PUBMED 9110172
 REFERENCE 3 (bases 1 to 251124)
 AUTHORS Boyesen, C., Inyoub, L., Smith, T.M., Smit, A., Wang, K., Rowen, L. and
 Hood, L.
 TITLE T-cell Receptor Alpha Delta Locus Complete Nucleotide Sequence
 JOURNAL Unpublished
 REFERENCE 4 (bases 1 to 251124)
 AUTHORS Boyesen, C., Lee, L., Smith, T.M., Smit, A., Kai, W., Lee, R. and Leroy, H.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-1997) Department of Molecular Biotechnology,
 University of Washington, Box 357330, Seattle, Washington 98195,
 USA
 REMARK Complete nucleotide sequence of the human T-cell receptor alpha

delta locus
On Jun 12, 2001 this sequence version replaced gi:2121231.
Sequencing method: high redundancy shotgun. Interspersed Repeats
were identified with RepeatMasker (available from
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>) Simple
sequence repeats were identified with sputnik (available from
<http://seraf.mbt.washington.edu/~chrisa/software/sputnik.html>).
Location/Qualifiers
1. 251124
/organism="Homo sapiens"
/db_xref="taxon:9606"
157356. .235794
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_11b="Mel Simon's BAC library"
/note="This span of sequence was submitted independently
as Genbank Accession Number U85197."
23572. .>251124
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BAC 378"
/clone_11b="Mel Simon's BAC library"
/note="This span of sequence was submitted independently
as Genbank Accession Number U85196."
/rpt_family="MIR"
603. 983
/rpt_family="MLT1B"
1738. .2478
/gene="TCRAV17S1"
1738. .1757
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="Old name TCRAV3S1"
join(<1957. .2008,2155. .>2439)
/gene="TCRAV17S1"
/product="TCRAV17S1"
/note="Old name TCRAV3S1"
join(1957. .2008,2155. .>2439)
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="Old name TCRAV3S1"
/codon_start=1
/product="TCRAV17S1"
/protein_id="AAB69018.1"
/db_xref="GI:2358043"
/translation="MTLLGVSLVILMLQLARVNSQGEEDPOLSTIOEGENATMNC
YKTSINNLQWYRQNSGRGLVHLILIRSNREKSGRLVTLDTSKSSILLITVASRA
DTASFCQAD"
join(1957. .2008,2155. .>2439)
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="Old name TCRAV3S1"
2440. .2446
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="heptamer"
2447. .2469
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="spacer"
2470. .2478
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="nonamer"
/rpt_family="MIR2"
4975. .5275
/rpt_family="AluY"
complement(5366. .5574)
/rpt_family="MERS8A"
7244. .7983
gene

/gene="TCRAV18S1"
/pseudo
7244. .7263
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/pseudo
join(<7482. .7527,7656. .>7944)
/gene="TCRAV18S1"
/product="TCRAV18S1"
/pseudo
join(7482. .7527,7656. .>7944)
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/pseudo
7945. .7951
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/note="heptamer"
7952. .7974
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/note="spacer"
7975. .7983
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/note="nonamer"
complement(8236. .8389)
/rpt_family="MER69A"
complement(8556. .8740)
/rpt_family="L1MB7"
complement(9281. .10170)
/rpt_family="L1PA2"
complement(10023. .11230)
/rpt_family="L1"
11234. .11405
/rpt_family="L1"
11818. .12491
/gene="TCRAV19S1"
11818. .11837
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="Old name TCRAV12S1"
join(<11900. .11948,12152. .>12452)
/gene="TCRAV19S1"
/product="TCRAV19S1"
/note="Old name TCRAV12S1"
join(11900. .11948,12152. .>12452)
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="Old name TCRAV12S1"
/codon_start=1
/product="TCRAV19S1"
/protein_id="AAB69019.1"
/db_xref="GI:2358044"
/translation="MLTASLRLAVIASICVSSMAQKVTQACTEISVEREKDYTLDCV
YETRDITLYLFWYKOPPSGEIVFLIRNSPDEQNEISGRYSNMFOKSTSFNFTLAS
QVDSAVYFCALSEA"
join(11900. .11948,12152. .>12452)
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="Old name TCRAV12S1"
12453. .12459
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="heptamer"
12460. .12482
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="spacer"
12483. .12491
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="nonamer"

```

repeat_region      13763..13833
                    /rpt_family="MIR2"
repeat_region      13879..13946
                    /rpt_family="MIR"
repeat_region      complement(13982..19677)
                    /rpt_family="MIR"
repeat_region      21487..21622
                    /rpt_family="MIR"
repeat_region      23034..23706
                    /rpt_family="MIR"
repeat_region      23955..24354
                    /rpt_family="MIR"
repeat_region      complement(24364..24523)
                    /rpt_family="MIR"
repeat_region      complement(24937..25045)

Query Match      83.0%; Score 415.6; DB 9; Length 251124;
Best Local Similarity 90.5%; Pred. No. 4.9e-123;
Matches 458; Conservative 0; Mismatches 39; Indels 9; Gaps 1;

4 TCGGCAACCTCCCAAGACGAGCTAGGTTTCTTGTGAGTGGGGAGTGAAGACAG 63
Db 247262 TCAGCAACCTCCCAAGACGAGCTAGGTTTCTTGTGAGTGGGGAGTGAAGACAG 247321
QY 64 GACTAGCTGATTTCTTGTGAGTGAAGATCCCTAAGCTAGCTGGAGTGAAGAC 123
Db 247322 GACTAGCTGATTTCTTGTGAGTGAAGATCCCTAAGCTAGCTGGAGTGAAGAC 247381
QY 124 ATCCACCTTTAAACGGGGCTTGCACCTTACCTACACCTGACCAATC-----AG 174
Db 247382 TTCCACCTTTAAACGGGGCTTGCACCTTACCTACACCTGACCAATCAGTGAAG 247441
QY 175 AGAGTCACTAAATGCTATATAGCAAAAGACAGAGTGAAGATAGCCATATCTA 234
Db 247442 AGAGTCACTAAATGCTATATAGCTAAAGACAGAGTGAAGATAGCCATATCTA 247501
QY 235 TTGCTGAGAGACAGACAGAGGAGCAATGATGGGATATTAACCCCAAGCTTTCAGCCG 294
Db 247502 TCACCTGAGAGACAGAGGAGGAGCAATGATGAGGATATTAACCCCAAGCTTTCAGCCG 247561
QY 295 GCAAGGCAACCCCTTTGGGTCCTCCCTTTGTATGGGAGCTCTGTTTATCTAT 354
Db 247562 GCMAAGGCAACCCCTTTGGGTCCTCCCTTTGTATGGGAGCTCTGTTTATCTAT 247621
QY 355 TCACCTATTAATCTTTCGACCTGCTTGTGTCATGTTTCTTTCAGGCTTTCAGCTG 414
Db 247622 TCACCTATTAATCTTTCGACCTGCTTGTGTCATGTTTCTTTCAGGCTTTCAGCTG 247681
QY 415 AGCTTTCGTCGCAATCCACACTGCTGTTTCCGCAACCGAGACCCGCTGACTCC 474
Db 247682 AGCTTTCGTCGCAATCCACACTGCTGTTTCCGCAATCCGCTGACTCC 247741
QY 475 CATCCCTGATCATGCAAGGCTGTC 500
Db 247742 CATCCCTGATCATGCAAGGCTGTC 247767

RESULT 13
AC092510/c 187321 bp DNA linear HTG 03-AUG-2002
LOCUS Papio cynocephalus anubis clone RP41-107F15, WORKING DRAFT
DEFINITION SEQUENCE, 5 ordered pieces.
ACCESSION AC092510
VERSION AC092510.2 GI:22094320
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE olive baboon.
ORGANISM Papio cynocephalus anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Papio.
REFERENCE 1 (bases 1 to 187321)
AUTHORS Akhter,N., Antonellis,A., Ayele,K., Eckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Hansen,N., Ho,S.-L., Ido,J.R., Karlins,E., Laric,P.,
Lee-Lin,S.-O., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
Margulies,E.H., Masello,C., Maekari,B., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pagutigan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Scheller,M.G., Sankarop,S., Thomas,J.W.,
Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggin,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 187321)
Green,E.D.
Direct Submission
Submitted (13-JUL-2001) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Galtersburg, MD 20877, USA
3 (bases 1 to 187321)
Green,E.D.
Direct Submission
Submitted (03-AUG-2002) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Galtersburg, MD 20877, USA
On Aug 3, 2002 this sequence version replaced gi:14717323.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@ngi.nih.gov
----- Project Information
Center project name: amf
Center clone name: 107F15

The sequence data in this record represents an 'enhanced'
version of a Phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8x average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 185897 bases at least Q40
Consensus quality: 185555 bases at least Q30
Consensus quality: 186768 bases at least Q20
Insert size: 165000; agarose-gel
Insert size: 149000; pulse-field-gel
Insert size: 186921; sum-of-contigs
Quality coverage: 12.32x in Q20 bases; agarose-gel
Quality coverage: 13.64x in Q20 bases; pulse-field-gel
Quality coverage: 10.87x in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
consists of 5 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 25945: contig of 25945 bp in length
25946 26045: gap of unknown length
26046 59962: contig of 33917 bp in length
59963 60062: gap of unknown length
60063 77104: contig of 17042 bp in length
77105 77204: gap of unknown length
77205 160478: contig of 83274 bp in length
160479 160578: gap of unknown length


```
family="L1"
```

244 AGCACAGCAGGAGGACATGATCGGATATAAACCCAACTTTCGAGCCGGCAACGGCA 303

Db 16276 AGCAGACAGAGGAGCAAGATTGGATATATATCCAGGATTCGAGCTGGCAACAGCA 16335
QY 304 ACCCGCTTGGGTCCTCCCTCTTGTATGGAGGCTGTGTTTCAATGATTTCACTAT 363
Db 16336 ACCCGCTTGGGTCCTCCCTCTTGTATGGAGG--CTGTTTCACTATTTCACTAT 16393
QY 364 TAAATCTGCAACTGCACTCTTGTGTCATGTTCTTACGCGCTTGAGCTGAGCTTTCGC 423
Db 16394 TAAATCTGCAACTGCACTCTTGTGTCATGTTTGTACTGTTGAGCTGACACTTTCAC 16453
QY 424 TCGCATTCACCACTGCTG-TTTGCCGCCGACCGAGACCCGCGCTGACTCCATCCCTC 482
Db 16454 TCGCATTCACCACTGCTGTTTGCCGCCGAGACCCGCTGACTCCATCCCTC 16513
QY 483 TGGATCAGCAGGGTTC 501
Db 16514 TGGATCAGCAGGGTTC 16532

RESULT 15
AL139038 140756 bp DNA linear PRI 15-JUL-2001
LOCUS DEFINITION Human DNA sequence from clone RP11-456B18 on chromosome 13,
complete sequence.
ACCESSION AL139038
VERSION AL139038.18 GI:14800148
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 140756)
Dunn, M.
Direct Submission
Submitted (14-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jul 17, 2001 this sequence version replaced gi:13751261.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-456B18 is from the library RPCI-11.2 constructed by the group
of Pletier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-456B18. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-456B18 is at 140756 in this
sequence. The true left end of clone RP11-78L16 is at 64441 in this
sequence. The true right end of clone RP11-108H9 is at 100 in this
sequence.
location/Qualifiers

source
1. 140756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-456B18"
/clone_11b="RPCI-11.2"
1830. .2326
/note="LOR1a repeat: matches 1. .497 of consensus"
3932. .3965
/note="17 copies 2 mer gt 94% conserved"
4632. .5067
/note="LOR1b repeat: matches 1. .461 of consensus"
5733. .5893
/note="L1ME3 repeat: matches 6002. .6146 of consensus"
6000. .6083
/note="MER34 repeat: matches 454. .539 of consensus"
6108. .6171
/note="HERV23 repeat: matches 993. .1056 of consensus"
7143. .7840
/note="LTR8 repeat: matches 1. .691 of consensus"
8059. .8292
/note="LTR16A repeat: matches 203. .442 of consensus"
8491. .8686
/note="LTR29 repeat: matches 6. .203 of consensus"
8725. .8926
/note="MER4D repeat: matches 483. .677 of consensus"
8992. .9049
/note="MER4D repeat: matches 410. .465 of consensus"
9062. .9281
/note="110 copies 2 mer tt 59% conserved"
9398. .9530
/note="MER41C repeat: matches 1. .131 of consensus"
9623. .9732
/note="LTR1 repeat: matches 675. .785 of consensus"
9970. .10333
/note="LTR2D repeat: matches 1. .486 of consensus"
10400. .10545
/note="LTR29 repeat: matches 2. .151 of consensus"
10549. .10935
/note="MER67D repeat: matches 12. .391 of consensus"
11249. .11332
/note="WIR repeat: matches 69. .144 of consensus"
11923. .12159
/note="MER4D repeat: matches 1. .234 of consensus"
12164. .12767
/note="MER4D repeat: matches 396. .973 of consensus"
12852. .13221
/note="THE1B repeat: matches 1. .364 of consensus"
13222. .14798
/note="THE1B-INTERNAL repeat: matches 1. .1580 of
consensus"
14799. .15173
/note="THE1B repeat: matches 1. .364 of consensus"
16144. .16346
/note="MER54B repeat: matches 585. .789 of consensus"
16339. .16503
/note="MER54B repeat: matches 112. .278 of consensus"
16504. .16571
/note="MER66A repeat: matches 1. .69 of consensus"
16570. .16629
/note="MER41B repeat: matches 283. .341 of consensus"
16599. .16902
/note="MER66A repeat: matches 163. .478 of consensus"
16903. .17014
/note="MER54B repeat: matches 3. .115 of consensus"
17195. .17491
/note="AluY repeat: matches 1. .297 of consensus"
18120. .19291
/note="L1P repeat: matches 1726. .2897 of consensus"
19460. .19584
/note="25 copies 5 mer attc 62% conserved"
19686. .19745
/note="12 copies 5 mer tttaa 71% conserved"

```
repeat_region 20453..20914
/note="LIMD2 repeat: matches 5745..6199 of consensus"
repeat_region 20913..21222
/note="LIMD2 repeat: matches 5277..5594 of consensus"
repeat_region 21308..21570
/note="LTR9 repeat: matches 357..625 of consensus"
repeat_region 21645..21984
/note="LTR19B repeat: matches 55..393 of consensus"
repeat_region 22081..22583
/note="LIMD2 repeat: matches 4726..5242 of consensus"
repeat_region 23498..23784
/note="AluDb repeat: matches 1..305 of consensus"
repeat_region 23873..24169
/note="AluSg repeat: matches 1..295 of consensus"
repeat_region 25044..25238
/note="MIR repeat: matches 31..238 of consensus"
repeat_region 25404..25696
/note="AluSx repeat: matches 20..311 of consensus"
repeat_region 25832..26218
/note="MSTC repeat: matches 1..405 of consensus"
repeat_region 26221..27865
/note="MSTC-internal repeat: matches 2..1651 of consensus"
repeat_region 27907..28659
/note="LIPB2 repeat: matches 5400..6155 of consensus"
repeat_region 28710..29043
/note="MSTD repeat: matches 30..394 of consensus"
repeat_region 29634..29697
/note="32 copies 2 mer ta 79% conserved"
repeat_region 29729..29752
/note="12 copies 2 mer tg 95% conserved"
repeat_region 30688..30874
/note="MER61A repeat: matches 160..354 of consensus"
repeat_region 31648..31785
/note="LIMC4 repeat: matches 6409..6540 of consensus"
repeat_region 31813..32243
/note="LIMC4 repeat: matches 6629..7089 of consensus"
repeat_region 32449..32821
/note="THEIC repeat: matches 1..371 of consensus"
repeat_region 32824..34464
/note="THEIC-internal repeat: matches 3..1651 of consensus"
repeat_region 34465..34825
/note="THEIC repeat: matches 1..360 of consensus"
repeat_region 34804..35004
/note="MER67C repeat: matches 202..403 of consensus"
repeat_region 39792..39868
/note="LIPB12 repeat: matches 6076..6152 of consensus"
repeat_region 40953..41258
/note="AluY repeat: matches 1..308 of consensus"
repeat_region 41279..41308
/note="15 copies 2 mer aa {6% conserved"
repeat_region 42744..42933
/note="LIMB8 repeat: matches 6064..6256 of consensus"
repeat_region 42973..43275
/note="AluSx repeat: matches 1..303 of consensus"
repeat_region 43560..43991
/note="match: STS: Em:HSPH07E5"
repeat_region 45196..45584
/note="MIR2B repeat: matches 2..399 of consensus"
repeat_region 45594..45767
/note="87 copies 2 mer ta 77% conserved"
repeat_region 45800..45851
/note="MIR2B repeat: matches 394..448 of consensus"
repeat_region 48118..48487
/note="THEIC repeat: matches 1..370 of consensus"
repeat_region 48708..48806
/note="AluO/FRAM repeat: matches 197..297 of consensus"
repeat_region 49853..50271
/note="LIMC repeat: matches 2274..2369 of consensus"
repeat_region 50637..50927
/note="LIME repeat: matches 974..1253 of consensus"
repeat_region 52250..52349
/note="MIR repeat: matches 131..228 of consensus"
```

```
repeat_region 52696..52996
/note="AluSg repeat: matches 1..293 of consensus"
repeat_region 54927..55010

Query Match 81.3% Score 407.2; DB 9; Length 140756;
Best Local Similarity 91.9%; Pred. No. 2,4e-120;
Matches 441; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 4 TCGGCAACCTCCCAACAGACACTTAACTTCTCTGTGAGATGGGGACTGAGACAG 63
Db 69419 TTGGCAACTTCCCAACAGACTTGGGTTTCTCTGTGAGAGGGGAGTGAAGACAG 69478

QY 64 GACTAGCTGATTTCTTCAAGGCTGACTAAGATCCCTTAAGCTTACTGGAAGTGACAC 123
Db 69479 GACTAGTTGATTTCTTCAAGGCACTAAGATCCCTTAAGCTTACTGGAAGTGACCG 69538

QY 124 ATCCACCTTTAAACAGGGGCTTCAACTTATGCTCACCTGACCAATCAGAGAGCTAC 183
Db 69539 ATCCACCTTTAAACATGGGGCTTGCACCTTACTACACCCACCAATCAGAGAGCTAC 69598

QY 184 TAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCAATCATATTGCTGAG 243
Db 69599 TAAATGCTAATTAGGCAAAAGAGAGGTAAAGAAATAGCCAATCATATTGCTGAA 69658

QY 244 AGCAGCAGAGAGGACATGATCGGATATAAACCCAGTCTTGAGCCGGCAACGGCA 303
Db 69659 AGCAGCAGAGAGGACATGATCGGATATAAACCCAGATTCGAGCCGGCAACGCA 69718

QY 304 ACCCCCTTTGGGTCCTCCCTCTTTGATGGAGCTGTGTTTCATGCTATTCTACTAT 363
Db 69719 ACCCCCTTTGGGTCCTCCCTCTTTGATGGAGCTGTGTTTCATGCTATTCTACTAT 69778

QY 364 TAAATCTGCAACTGCACTCTTCTGTCATGTTCTTTCAGGCTTGAAGCTGAGC-TTTCG 422
Db 69779 TAAATCTGCAACTGCACTCTTCTGTCATGTTCTTTCAGGCTTGAAGCTGAGCTTTTCG 69838

QY 423 CTCGCATCCACCACTGCTGTTTGGCCGACCGACGAGACCCCGCTGATCCCATCCCTC 482
Db 69839 CTCGCCTGCCACCACTGCTGTTTGGCACTGTGCGAGACCCCGCTGATCCCATCCCGC 69898
```

Search completed: April 19, 2003, 13:59:12
Job time : 2138.47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 10:05:38 ; Search time 180.12 Seconds
(without alignments)
6263.892 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501
Sequence: 1 gtcctgcacacactcccca.....ctgcatcagcaggtctcc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	10499	21	ABN97929
2	498	99.4	56093	24	ABL61744
3	461.4	92.1	7466	23	AA568626
4	450	89.8	711	21	ABN97947
5	426	85.0	1329	19	AAV43219
6	426	85.0	1329	20	AAZ29704
7	425.6	85.0	783	20	AAZ5669
8	396.2	79.1	711	21	ABN97946
9	396.2	79.1	1393	22	AA531002

10	377.8	75.4	893	23	AA565964	DNA encoding novel
11	377	75.2	2781	22	AA555630	Nucleotide sequenc
12	375.4	74.9	2946	20	AA775526	Human secreted pro
13	375.4	74.9	2946	21	AA59468	Human secreted pro
14	374.4	74.7	2930	24	AA24195	Human syncytin CDN
15	373.4	74.5	7582	20	AA25565	Complete human end
16	373.4	74.5	7582	21	AA59215	Human endogenous r
17	372.6	74.4	849	22	AA51000	Human diagnostic a
18	372.2	74.3	3372	20	AA25663	Human endogenous r
19	372.2	74.3	3372	21	AA59213	Partial pol gene a
20	371.2	74.1	1136	20	AA25660	Human endogenous r
21	371.2	74.1	1136	21	AA59210	Human endogenous r
22	371.2	74.1	1136	22	AA59217	3' pol gene and 3'
23	371.2	74.1	1136	23	AA59217	DNA encoding novel
24	371.2	74.1	1136	24	AA59217	DNA encoding novel
25	371.2	74.1	1136	25	AA59217	DNA encoding novel
26	370.6	74.0	2782	20	AA25661	DNA encoding novel
27	370.6	74.0	2782	21	AA59211	Human endogenous r
28	370.6	74.0	2782	22	AA20069	5' non coding, 3'
29	350.8	70.0	2300	23	AA59268	HERV-W envelope pr
30	347.4	69.3	1165	23	AA572721	DNA encoding novel
31	346	69.1	2942	23	AA577313	DNA encoding novel
32	345.8	69.0	2629	22	AA577491	DNA encoding novel
33	341.8	68.2	635	19	AAV43215	Human immune/haema
34	341.8	68.2	635	20	AA529702	Multiple sclerosis
35	341.8	68.2	2030	21	AA53826	Nucleotide sequenc
36	325.6	65.0	808	23	AA588392	DNA encoding novel
37	321.8	64.2	1478	23	AA587568	DNA encoding novel
38	320.8	64.0	1243	23	AA584189	DNA encoding novel
39	312.6	62.4	448	23	AA587567	DNA encoding novel
40	312.6	62.4	583	23	AA588391	DNA encoding novel
41	311.4	62.2	410	24	ABN94040	Gene #538 used to
42	311.4	62.2	410	24	ABK64210	Human benign prost
43	308.2	61.5	1758	22	AAK64138	Human immune/haema
44	303.4	60.6	439	22	AA53993	Human polynucleoti
45	303.4	60.6	439	22	AA531671	Genomic sequence #

ALIGNMENTS

RESULT 1	ABN97929	standard; DNA; 10499 BP.
ID	ABN97929	
AC	ABN97929	
XX		
DT	01-AUG-2002	(first entry)
DE	Human retroviral sequence HERV-7q.	
XX		
KW	Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;	
KW	multiple sclerosis; ds.	
XX		
OS	Human retrovirus.	
XX		
PN	WO9967395-A1.	
XX		
PD	29-DEC-1999.	
XX		
PF	23-JUN-1999;	99WO-FR01513.
XX		
PR	23-JUN-1998;	98FR-0007920.
XX		
PA	(INRM) INSEPM INST NAT SANTB & RECH MEDICALE.	
XX		
PI	Alliel PM, Perin J, Rieger F;	
XX		
DR	WPI; 2000-160587/14.	
XX		
FT	New nucleic acid sequences of human endogenous retrovirus, HERV-7q,	
PT	used for diagnosis, treatment and prevention of autoimmune and	
PT	neurological diseases -	

XX Claim 3; Fig 1; 225pp; French.
XX
XX The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;
Query Match 100.0%; Score 501; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 8,2e-161;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCTGCGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGA 60
Db 9500 GTCTGCGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGA 9559
QY 61 CAGAGCTAGCTGAGATTTCTTCTAGCTGAGATGATCCCTAAGCTGAGGAGGAGTAC 120
Db 9560 CAGAGCTAGCTGAGATTTCTTCTAGCTGAGATGATCCCTAAGCTGAGGAGGAGTAC 9619
QY 121 CACATCCACCTTTAAACAAGGGGCTTGCACTTAGTTCACACCTGACCAATCAGAGAGCT 180
Db 9620 CACATCCACCTTTAAACAAGGGGCTTGCACTTAGTTCACACCTGACCAATCAGAGAGCT 9679
QY 181 CACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCATCTATTGCTT 240
Db 9680 CACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCATCTATTGCTT 9739
QY 241 GAGACACAGCAGGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCGCGCAACG 300
Db 9740 GAGACACAGCAGGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCGCGCAACG 9799
QY 301 GCAACCCCTTTGGGCTCCCTCTTTGATGGAGTCTGTGTTTCAATGCTATTTCACCTC 360
Db 9800 GCAACCCCTTTGGGCTCCCTCTTTGATGGAGTCTGTGTTTCAATGCTATTTCACCTC 9859
QY 361 TATTAAATCTTGCAACTGACTCTTCTGTGCTCATGTTCTTAAAGGCTTGAGCTTTC 420
Db 9860 TATTAAATCTTGCAACTGACTCTTCTGTGCTCATGTTCTTAAAGGCTTGAGCTTTC 9919
QY 421 CGCTGCGCATCCACAGCAGTGTGCGCGCAACGCAACCGCGCGCTGACCTCCATCCC 480
Db 9920 CGCTGCGCATCCACAGCAGTGTGCGCGCAACGCAACCGCGCGCTGACCTCCATCCC 9979
QY 481 TCTGGATCATGAGGAGTGTCC 501
Db 9980 TCTGGATCATGAGGAGTGTCC 10000

KM gene; ds.
XX
XX
OS Homo sapiens.
XX
XX WO200194629-A2.
XX
XX 13-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-US10838.
XX
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-231133P.
XX 18-SEP-2000; 2000US-231617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 20-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 25-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235863P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.
XX 28-SEP-2000; 2000US-236109P.
XX 28-SEP-2000; 2000US-236111P.
XX 29-SEP-2000; 2000US-236842P.
XX 29-SEP-2000; 2000US-236891P.
XX 02-OCT-2000; 2000US-237172P.
XX 02-OCT-2000; 2000US-237173P.
XX 02-OCT-2000; 2000US-237278P.
XX 02-OCT-2000; 2000US-237294P.
XX 02-OCT-2000; 2000US-237295P.
XX 02-OCT-2000; 2000US-237316P.
XX 03-OCT-2000; 2000US-237425P.
XX 03-OCT-2000; 2000US-237598P.
XX 03-OCT-2000; 2000US-237604P.
XX 03-OCT-2000; 2000US-237606P.
XX 03-OCT-2000; 2000US-237608P.
XX 01-NOV-2000; 2000US-244867P.
XX 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Sopet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 81; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic

CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.

XX Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16861 T; 0 other;

Query Match 99.4%; Score 498; DB 24; Length 56093;

Best Local Similarity 100.0%; Pred. No. 2,2e-159;

Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63

Db 37504 TCGGCCAACCTCCCAACAGCACTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 37563

Qy 64 GACTAGCTGAGATTCTTCTAGGCTGATTAAGATCCCTAAGCCTAGCTGGGAAGTACAC 123

Db 37564 GACTAGCTGAGATTCTTCTAGGCTGATTAAGATCCCTAAGCCTAGCTGGGAAGTACAC 37623

Qy 124 ATCCACCTTTAAACAGGGGGCTTGCACTTACCTACACCTGACCAATCAGAGAGCTCAC 183

Db 37624 ATCCACCTTTAAACAGGGGGCTTGCACTTACCTACACCTGACCAATCAGAGAGCTCAC 37683

Qy 184 TAAATGCTAATTAGGCAAGAAGAGAGTAAAGAAATGCAATCATCTATTGCTTGAG 243

Db 37684 TAAATGCTAATTAGGCAAGAAGAGAGTAAAGAAATGCAATCATCTATTGCTTGAG 37743

Qy 244 AGCAGACAGAGAGGAGCAATGATCGGATATTAACCCAGTCTTGAGAGCGGCAACGGCA 303

Db 37744 AGCAGACAGAGAGGAGCAATGATCGGATATTAACCCAGTCTTGAGAGCGGCAACGGCA 37803

Qy 304 ACCCCCTTTGGGTCCTCCCTCTTTGATGGAGCTCTGTTTCATGCTATTTCACCTTAT 363

Db 37804 ACCCCCTTTGGGTCCTCCCTCTTTGATGGAGCTCTGTTTCATGCTATTTCACCTTAT 37863

Qy 364 TAAATCTTGCACTGCACTCTTCTGTCATGTTTCTTAACGGCTTGAGCTGAGCTTTCGC 423

Db 37864 TAAATCTTGCACTGCACTCTTCTGTCATGTTTCTTAACGGCTTGAGCTGAGCTTTCGC 37923

Qy 424 TCGCCATCCACCACTGCTGTTTGGCCACCGGCAACCGCGCTGACTCCCATCCCTCT 483

Db 37924 TCGCCATCCACCACTGCTGTTTGGCCACCGGCAACCGCGCTGACTCCCATCCCTCT 37983

Qy 484 GGATCATGAGGGGTGCC 501

Db 37984 GGATCATGAGGGGTGCC 38001

RESULT 3

AAS68626 ID AAS68626 standard; cDNA, 7466 BP.

XX AAS68626;

DT 13-FEB-2002 (first entry)

DB DNA encoding novel human diagnostic protein #4430.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG04439.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX Claim 1; SEQ ID No 4430; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridization probes,

XX and gene mapping, and in recombinant production of (II). The

XX polynucleotides are also used in diagnostics as expressed sequence tags

XX for identifying expressed genes. (I) is useful in gene therapy techniques

XX to restore normal activity of (II) or to treat disease states involving

XX (II). (II) is useful for generating antibodies against it, detecting or

XX quantitating a polypeptide in tissue, as molecular weight markers and as

XX a food supplement. (II) and its binding partners are useful in medical

XX imaging of sites expressing (II). (I) and (II) are useful for treating

XX disorders involving aberrant protein expression or biological activity.

XX The polypeptide and polynucleotide sequences have applications in

XX diagnostics, forensics, gene mapping, identification of mutations

XX and to produce other types of data and products dependent on DNA and

XX amino acid sequences. AAS64197-AAS94564 represent novel human

XX diagnostic coding sequences of the invention.

XX Note: The sequence data for this patent did not appear in the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;

Qy Query Match 92.1%; Score 461.4; DB 23; Length 7466;

Best Local Similarity 96.6%; Pred. No. 2.8e-147;

Matches 482; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

Qy 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63

Db 3668 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 37227

Qy 64 GACTAGCTGAGATTCTTCTAGGCTGATTAAGATCCCTAAGCCTAGCTGGGAAGTACAC 123

Db 3728 GACTAGCTGAGATTCTTCTAGGCTGATTAAGATCCCTAAGCCTAGCTGGGAAGTACAC 3787

Qy 124 ATCCACCTTTAAACAGGGGGCTTGCACTTACCTACACCTGACCAATCAGAGAGCTCAC 183

Db 3788 ATCCACCTTTAAACAGGGGGCTTGCACTTACCTACACCTGACCAATCAGAGAGCTCAC 3847

Qy 184 TAAATGCTAATTAGGCAAGAAGAGAGTAAAGAAATGCAATCATCTATTGCTTGAG 243

Db 3848 TAAATGCTAATTAGGCAAGAAGAGAGTAAAGAAATGCAATCATCTATTGCTTGAG 3907

Qy 244 AGCAGACAGAGAGGAGCAATGATCGGATATTAACCCAGTCTTGAGAGCGGCAACGGCA 303

Db 3908 AGCAGACAGAGAGGAGCAATGATCGGATATTAACCCAGTCTTGAGAGCGGCAACGGCA 3967

Qy 304 -ACCCCTTTGGGTCCTCCCTCTTTGATGGAGCTCTGTTTCATGCTATTTCACCTTA 362

Db 3968 GCCCCTTTGGGTCCTCCCTCTTTGATGGAGCTCTGTTTCATGCTATTTCACCTTA 4027

Qy 363 TTAATCTTGCACTGCACTCTTCTGTCATGTTTCTTAACGGCTTGAGCTGAGCTTTCG 422

Db 4028 TTAATCTTCACTGACTCTTCTGTGTCATGTTTCTTACGGCTCGAGCTGAGCTTTTG 4087
 Qy 423 CTGGCATCACCACCTGCTGTTTGCGGCAACGGACAGCCGGCGTGACTCCCATCCCTC 482
 Db 4088 CTCACGCTCACCACCTGCTGTTTGCCACCAACGGACAGCTGCGCTGACTCCCATCCCTC 4147
 Qy 483 TGGATCATGCAAGGTGTCC 501
 Db 4148 TGGATCTCGCAGAGGTGTCC 4166

RESULT 4

ABN97947
 ID ABN97947 standard; DNA; 711 BP.

XX AC ABN97947;

XX DT 01-AUG-2002 (first entry)

XX DE Human retroviral sequence R1F.

XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 KM multiple sclerosis; ds.

XX OS Human retrovirus.

XX PN MO967395-A1.

XX PD 29-DEC-1999.

XX PF 23-JUN-1999; 99WO-FR01513.

XX PR 23-JUN-1998; 98FR-0007920.

XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX PI Allele PM, Perin J, Rieger F;

XX DR WPI; 2000-160587/14.

XX PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
 PT used for diagnosis, treatment and prevention of autoimmune and
 PT neurological diseases -

XX PS Claim 3; Fig 3; 225pp; French.

XX CC The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer.
 CC neuropeptides or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention.

XX CC Sequence 711 BP; 171 A; 209 C; 160 G; 171 T; 0 other;

XX SQ

Query Match 89.8%; Score 450; D3 21; Length 711;
 Best Local Similarity 100.0%; Pred. No. 6; 4e-144;
 Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 ACTGAGAGACAGACTAGTGTGATTTCCTAGGCTGACTAAGAACTCCTAAGCTTACCTGG 111

Db 1 ACTGAGAGACAGACTAGTGTGATTTCCTAAGCTGACTAAGAACTCCTAAGCTTACCTGG 60

Qy 112 GAAGGTGACCACTCCACTTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAAT 171

Db 61 GAAGGTGACCACTCCACTTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAAT 120

Qy 172 CAGAGAGCTCACTAAATGCTAATTAGCAAAAGACAGAGGTAAAGAAATAGCCAATCAT 231
 Db 121 CAGAGAGCTCACTAAATGCTAATTAGCAAAAGACAGAGGTAAAGAAATAGCCAATCAT 180
 Qy 232 CTATTGCTTGAGAGACACGACGAGGGGCAATGATCGGGAATATAACCAAGCTCTTGAG 291
 Db 181 CTATTGCTTGAGAGACACGACGAGGGGCAATGATCGGGAATATAACCAAGCTCTTGAG 240
 Qy 292 CCGGCAACGCAACCCCTTTGGGTCCTCCCTTTGATGAGAGCTGTGTTTCATGCT 351
 Db 241 CCGGCAACGCAACCCCTTTGGGTCCTCCCTTTGATGAGAGCTGTGTTTCATGCT 300
 Qy 352 ATTTCACCTATTAATCTTTCGCACTGTCACCTCTTCTGCTCCATGTTTCTTAAGCTTGAG 411
 Db 301 ATTTCACCTATTAATCTTTCGCACTGTCACCTCTTCTGCTCCATGTTTCTTAAGCTTGAG 360
 Qy 412 CTGAGCTTTGCGTGGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGGCTGAC 471
 Db 361 CTGAGCTTTGCGTGGCCATCCACCACTGCTGTTTGCCGCCACCGCAGACCCGGCTGAC 420
 Qy 472 TCCCATCCCTCGATCATGACGAGGTGTCC 501
 Db 421 TCCCATCCCTCGATCATGACGAGGTGTCC 450

RESULT 5

AAV43219
 ID AAV43219 standard; CDNA; 1329 BP.

XX AC AAV43219;

XX DT 29-DEC-1998 (first entry)

XX DE Multiple sclerosis associated retrovirus fragment 6.

XX KW Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
 KW gag gene; env gene; rheumatoid arthritis-associated virus; ss.

XX OS Multiple sclerosis associated retrovirus.

XX PN MO9823755-A1.

XX PD 04-JUN-1998.

XX PF 26-NOV-1997; 97WO-1B01482.

XX PR 26-NOV-1996; 96US-0756429.

XX PA (INMR) BIO MERIEUX.

XX PI Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F;
 PI Mandrand B, Paranhos-Baccala G, Perron H;

XX DR WPI; 1998-322732/28.

XX PF P-PSDB; AAW71069.

XX PT New nucleic acid from retroviruses - useful for diagnosis,
 PT prevention and treatment of, e.g. multiple sclerosis

XX PS Disclosure; Pages 187-188; 286pp; English.

CC The present sequence represents a multiple sclerosis (MS) associated
 CC retrovirus (MSRV) genomic fragment used in the method of the
 CC invention. The invention provides complete or partial genomic
 CC sequences of the MSRV-1 pol gene, gag gene and env gene and
 CC polypeptides encoded by these genes. The invention also provides
 CC antibodies raised against the polypeptides. The genomic sequences,
 CC polypeptides and antibodies are also claimed useful for diagnosing
 CC infection by MS and rheumatoid arthritis-associated viruses, and also
 CC for prevention and treatment of infection with these viruses.

XX
 SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 85.0%; Score 426; DB 19; Length 1329;
 Best Local Similarity 90.6%; Pred. No. 1.6e-135;
 Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 2 TCTGGCCCACTCCCAACAGACACTTAGTTTCTGTTGAGATGGGGAGCTGAGAGAC 61
 Db TGTGAGCCCACTCCCAACAGACTTGGGTTTCTGTTGAGAGGGTGAAGAGAGAC 556
 Qy 62 AGAAGTACCTGATTTCTAGGCTGACTAAGATCCCTAAGCCTAGTGGAGAGGTGACC 121
 Db 557 AGGATAGCTGATTTCTAGGCTGACTAAGATCCCTAAGCCTAGTGGAGAGGTGACC 616
 Qy 122 ACATCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCAACCTGACCAATCAGAGAGCTC 181
 Db 617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCAACCCGACCAATCAGAGAGCTC 676
 Qy 182 ACTAAATGCTAATTAAGCAAGAGAGTAAAGAAATAGCCATATCTATTGCTCG 241
 Db 677 ACTAAATGCTAATCAAGCAAGAGAGTAAAGCAATAGCCATATCTATTGCTCG 736
 Qy 242 AGAGCAGCAGAGAGAGGACAATGATCGGGATATAAACCCAACTTTCGAGCCGCAACGG 301
 Db 737 AGAGCAGCAGCGGAGAGAGAGCAAGATTTGGATATAACTAGCATTACAGCCAGAACAG 796
 Qy 302 CAACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTCTGTTTCAATGCTATTACTCT 361
 Db 797 CAACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTCTGTTTCAATGCTATTACTCT 856
 Qy 362 ATTAATCTTGCACTGCACTTTCTGTCATGTTTCTTACGCGTTGAGCTGAGCTTTC 421
 Db 857 ATTAATCTTGCACTGCACTTTCTGTCATGTTTCTTATGCTCAAGCTGAGCTTTC 916
 Qy 422 GCTGCCATTCACCACTGCTGTTTGGCCGACCGGACCGCGCTGATCCCATCCCT 481
 Db 917 GTTCCCATTCACCACTGCTGTTTGGCCGACCGCTGATCCCATCCCT 976
 Qy 482 CTGATCATGACGAGGTGTC 501
 Db 977 TTGATCCAGCAGAGTGTCC 996

RESULT 6
 AAX29704
 ID AAX29704 standard; DNA; 1329 BP.

XX AAX29704;
 AC 08-JUN-1999 (first entry)
 XX Clone 5M6 from MSRV-1.
 DE Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
 KM rheumatoid polyarthritits; ss.
 OS Multiple sclerosis related virus type 1.
 XX FR2765588-A1.
 XX 08-JAN-1999.
 PD 07-JUL-1997; 97FR-0008816.
 XX PF

XX 07-JUL-1997; 97FR-0008816.
 PR (INMR) BIO MERIEUX.
 XX WPI; 1999-098275/09.
 DR P-PSDB; AAW99554.
 XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
 PT multiple sclerosis or rheumatoid polyarthritits
 PS Claim 1; Page 39-40; 83pp; French.
 XX This sequence represents clone 5M6 from a novel multiple sclerosis
 CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
 CC prophylactic or therapeutic compositions to inhibit expression of a
 CC multiple sclerosis related virus and/or virus associated with
 CC rheumatoid polyarthritits.

Query Match 85.0%; Score 426; DB 20; Length 1329;
 Best Local Similarity 90.6%; Pred. No. 1.6e-135;
 Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 2 TCTGGCCCACTCCCAACAGACACTTAGTTTCTGTTGAGATGGGGAGCTGAGAGAC 61
 Db 497 TGTGAGCCCACTCCCAACAGACTTGGGTTTCTGTTGAGAGGGTGAAGAGAGAC 556
 Qy 62 AGAAGTACCTGATTTCTAGGCTGACTAAGATCCCTAAGCCTAGTGGAGAGGTGACC 121
 Db 557 AGGATAGCTGATTTCTAGGCTGACTAAGATCCCTAAGCCTAGTGGAGAGGTGACC 616
 Qy 122 ACATCCACCTTTAAACAGCGGGCTTGCAACTTAGCTCAACCTGACCAATCAGAGAGCTC 181
 Db 617 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCAACCCGACCAATCAGAGAGCTC 676
 Qy 182 ACTAAATGCTAATTAAGCAAGAGAGTAAAGAAATAGCCATATCTATTGCTCG 241
 Db 677 ACTAAATGCTAATCAGCAAGAGAGTAAAGCAATAGCCATATCTATTGCTCG 736
 Qy 242 AGAGCAGCAGAGAGAGGACAATGATCGGGATATAAACCCAACTTTCGAGCCGCAACGG 301
 Db 737 AGAGCAGCAGCGGAGAGAGCAAGATTTGGATATAACTAGCATTACAGCCAGAACAG 796
 Qy 302 CAACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTCTGTTTCAATGCTATTACTCT 361
 Db 797 CAACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTCTGTTTCAATGCTATTACTCT 856
 Qy 362 ATTAATCTTGCACTGCACTTTCTGTCATGTTTCTTACGCGTTGAGCTGAGCTTTC 421
 Db 857 ATTAATCTTGCACTGCACTTTCTGTCATGTTTCTTATGCTCAAGCTGAGCTTTC 916
 Qy 422 GCTGCCATTCACCACTGCTGTTTGGCCGACCGGACCGCGCTGATCCCATCCCT 481
 Db 917 GTTCCCATTCACCACTGCTGTTTGGCCGACCGCTGATCCCATCCCT 976
 Qy 482 CTGATCATGACGAGGTGTC 501
 Db 977 TTGATCCAGCAGAGTGTCC 996

RESULT 7
 AAX25669
 ID AAX25669 standard; cDNA to mRNA; 783 BP.

XX AAX25669;
 AC 21-MAY-1999 (first entry)
 XX Human endogenous retrovirus w long terminal repeat region.
 DE Clone; human endogenous retrovirus; genome; autoimmune disease;
 XX KW

KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
 OS Human endogenous retrovirus.
 XX WO9902696-A1.
 XX 21-JAN-1999.
 XX PD
 XX PF 06-JUL-1998; 98WO-FR01442.
 XX PR 07-JUL-1997; 97FR-0008615.
 XX PA (INMR) BIO MERIEUX.
 XX PI Beeseme F, Blond JL, Bouton O, Mallet F, Mandrand B;
 XX WPI; 1999-120897/10.
 XX DR
 XX PT New nucleic acid sequences from human endogenous retrovirus-W -
 PT expressed exclusively in placenta and useful in diagnosis and
 PT therapy of autoimmune disease, and abnormal or failed pregnancy
 XX
 XX Claim 1; Page 83; 106pp; French.
 XX
 XX This sequence represents the long terminal repeat (LTR) region of the
 CC human endogenous retrovirus (HERV) W genome. The nucleic acids, their
 CC fragments or peptides encoded by them are markers of autoimmune disease
 CC (e.g. multiple sclerosis, rheumatoid polyarthritis, disseminated lupus
 CC erythematosus), insulin-dependent diabetes and related pathologies) and
 CC of abnormal or unsuccessful pregnancy and can be used as chromosomal
 CC markers for susceptibility to these conditions, or proximity markers
 CC of genes associated with this susceptibility.
 XX
 XX Sequence 783 BP; 173 A; 213 C; 166 G; 180 T; 51 other;
 SQ
 Query Match 85.0%; Score 425.6; DB 20; Length 783;
 Best Local Similarity 94.0%; Pred. No. 1.6e-135;
 Matches 422; Conservative 25; Mismatches 1; Indels 1; Gaps 1;
 QY 54 TGAAGACGAGCTAGCTGATTTCTTAGCGACTAGAAATCCCTAAGCCCTAGGCGGA 113
 DB 1 TGAAGACGAGCTAGCTGATTTCTTAGCGACTAGAAATCCCTAAGCCCTAGGCGGA 60
 QY 114 AGGTGACCATCCACCTTTAAACAGGGGCTTGCACCTTAGCTCAGACCTGACCAATCA 173
 DB 61 AGGTGACCATCCACCTTTAAACAGGGGCTTGCACCTTAGCTCAGACCTGACCAATCA 120
 QY 174 GAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCCATCATCT 233
 DB 121 GAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCCATCATCT 180
 QY 234 ATTGCTGAGACAGAGGAGGAGCAATGATCCGGATTAATCCCAAGCTTGAGGCC 293
 DB 181 ATTGCTGAGACAGAGGAGGAGCAATGATCCGGATTAATCCCAAGCTTGAGGCC 240
 QY 294 GGCACAGGCA-ACCCCTTTGGTCCCTCCCTTTGTATGGGAGCTCTTTTCAATGCTA 352
 DB 241 GGCACAGGCAAGGCCCTTTGGTCCCTCCCTTTGTATGGGAGCTCTTTTCAATGCTA 300
 QY 353 TTTCACCTATTAAATCTTGCAACGTGCACTTTCTGTCATGTTTCTTACGCTTACG 412
 DB 301 TTTCACCTATTAAATCTTGCAACGTGCACTTTCTGTCATGTTTCTTACGCTTACG 360
 QY 413 TGAAGCTTTGGCTGCGCATCCACACGCTGTTTGGCCGACCCGAGACACCCGCGCTGACT 472
 DB 361 TGAAGCTTTGGCTGCGCATCCACACGCTGTTTGGCCGACCCGAGACACCCGCGCTGACT 420
 QY 473 CCATCCCTCTGGATCATGAGGGGTGCC 501
 DB 421 CCATCCCTCTGGATCATGAGGGGTGCC 449

RESULT 8
 ABN97946
 ID ABN97946 standard; DNA; 711 BP.
 XX
 XX AC ABN97946;
 XX
 XX DT 01-AUG-2002 (first entry)
 XX
 XX DE Human retroviral sequence R1.
 XX
 XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 KW multiple sclerosis; ds.
 XX
 XX OS Human retrovirus.
 XX PN WO9967395-A1.
 XX PD 29-DEC-1999.
 XX PF 23-JUN-1999; 99WO-FR01513.
 XX PR 23-JUN-1998; 98FR-0007920.
 XX PA (INMR) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 XX Alliel PM, Perin J, Rieger F;
 XX WPI; 2000-160587/14.
 XX DR
 XX PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
 PT used for diagnosis, treatment and prevention of autoimmune and
 PT neurological diseases -
 XX
 XX PS Claim 3; Fig 3; 225pp; French.
 XX
 XX The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention.
 XX
 XX Sequence 711 BP; 170 A; 204 C; 162 G; 175 T; 0 other;
 SQ
 Query Match 79.1%; Score 396.2; DB 21; Length 711;
 Best Local Similarity 95.3%; Pred. No. 2e-125;
 Matches 430; Conservative 0; Mismatches 18; Indels 3; Gaps 2;
 QY 52 ACTGAGACAGGACTAGCTGATTTCTTAGCGCTGACTAAGAAATCCCTAAGCTTACTGG 111
 DB 1 ACTGAGACAGGACTAGCTGATTTCTTAGCGCTGACTAAGAAATCCCTAAGCTTACTGG 60
 QY 112 GAAAGTGACACAGCTCACCCTTTAAACAGGGGCTTGCACCTTAGCTCACACCTGACCAAT 171
 DB 61 GAAAGTGACACAGCTCACCCTTTAAACAGGGGCTTGCACCTTAGCTCACACCTGACCAAT 120
 QY 172 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTTAAAGAAATAGCCATCAT 231
 DB 121 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTTAAAGAAATAGCCATCAT 180
 QY 232 CTATTGCTGAGACAGACAGAGGAGCAATGATCGGATTAATCCCAAGTCTTGCAG 291
 DB 181 CTATTGCTGAGACAGACAGAGGAGCAATGATCGGATTAATCCCAAGTCTTGCAG 240
 QY 292 CCGGCAAGGCA-ACCCCTTTGGTCCCTCCCTTTGTATGGGAGCTCTGTTTCAATG 350
 DB 240 CCGGCAAGGCA-ACCCCTTTGGTCCCTCCCTTTGTATGGGAGCTCTGTTTCAATG 350

Dh 241 CTGCAACAGACGCCCCCTTTGGTCCCTTCCCTTGTGATGGAG--CTGTTTTCATGC 298
Qy 351 TATTTCACTATTAAATCTTGCAACTGACCTTCTGGCCAGTTTCTTACGACTGA 410
Dh 299 TATTTCACTATTAAATCTTGCAACTGACCTTCTGGCCAGTTTCTTACGACTGA 358
Qy 411 GCTGAGCTTTGCTGCGCATCCACCACTGCTGTTTGGCCGACGACGCCGCGCTGA 470
Dh 359 GCTGAGCTTTGCTGCGCATCCACCACTGCTGTTTGGCCGACGACGCCGCGCTGA 418
Qy 471 CTCCATCCCTCTGATCATGACGAGGTCC 501
Dh 419 CTCCATCCCTCTGATCATGACGAGGTCC 449

RESULT 9

AAS31002 standard, cDNA; 1393 BP.

AAS31002;

04-DEC-2001 (first entry)

Human diagnostic and therapeutic polynucleotide (DITHP) #17.

KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder; ss.

Homo sapiens.

MO200162927-A2.

30-AUG-2001.

21-FEB-2001; 2001MO-US06059.

24-FEB-2000; 2000US-0184693.

24-FEB-2000; 2000US-0184697.

24-FEB-2000; 2000US-0184698.

24-FEB-2000; 2000US-0184769.

24-FEB-2000; 2000US-0184770.

24-FEB-2000; 2000US-0184771.

24-FEB-2000; 2000US-0184772.

24-FEB-2000; 2000US-0184773.

24-FEB-2000; 2000US-0184774.

24-FEB-2000; 2000US-0184776.

24-FEB-2000; 2000US-0184777.

24-FEB-2000; 2000US-0184797.

24-FEB-2000; 2000US-0184813.

24-FEB-2000; 2000US-0184837.

24-FEB-2000; 2000US-0184841.

24-FEB-2000; 2000US-0185213.

12-MAY-2000; 2000US-0185216.

12-MAY-2000; 2000US-0203785.

12-MAY-2000; 2000US-0204226.

12-MAY-2000; 2000US-0204255.

16-MAY-2000; 2000US-0204821.

16-MAY-2000; 2000US-0204908.

16-MAY-2000; 2000US-0205232.

16-MAY-2000; 2000US-0204815.

17-MAY-2000; 2000US-0204863.

17-MAY-2000; 2000US-0205221.

17-MAY-2000; 2000US-0205285.

17-MAY-2000; 2000US-0205286.

17-MAY-2000; 2000US-0205287.

17-MAY-2000; 2000US-0205323.

17-MAY-2000; 2000US-0205324.

XX (INCY-) INCYTE GENOMICS INC.

XX

PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'Sa SA, Ameh S, Dahl CR, Dam TC, Daniels SE;
PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockbreiter TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DW, Lincoln SE, Jackson S;
XX WPI; 2001-502867/55.
DR P-PSDB; AAU9431.
XX

Claim 1; Page 304; 522pp; English.

CC The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbent
CC assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and
CC therapeutic (DITHP) polynucleotides of the invention.

SQ Sequence 1393 BP; 375 A; 352 C; 304 G; 360 T; 2 other;

Query Match 79.1%; Score 396.2; DB 22; Length 1393;

Best Local Similarity 89.7%; Pred. No. 2.8e-125; Matches 455; Conservative 0; Mismatches 33; Indels 19; Gaps 2;

Qy 4 TCGGCAACCTCCCAACAGACACTTATGTTTCTGTTGAGATGGGGAGCTGAGACAG 63

Dh 197 TCAAGCAACTCCCAACAGACACTTATGTTTCTGTTGAGATGGGGAGCTGAGACAG 256

Qy 64 GACTAGCTGATTTCTAGGCTGATTAAGATCTTAAGCTTAAGCTGGAAGTGACAC 123

Dh 257 GACTAGCTGATTTCTAGGCTGATTAAGATCTTAAGCTTAAGCTGGAAGTGACAC 316

Qy 124 ATCCACCTTTAAACAGGGGCTTGAACCTTAAGCTTAAGCTTAAGCTTAAGCTTAAG 174

Dh 317 GTTAACTTTAAACAGGGGCTTGAACCTTAAGCTTAAGCTTAAGCTTAAGCTTAAG 376

Qy 175 AGAGCTCACTAAATGCTAATTAAGCAAGAGAGTAAAGTAAAGTAAAGTAAAGTAAAG 234

Dh 377 AGAGCTCACTAAATGCTAATTAAGCAAGAGAGTAAAGTAAAGTAAAGTAAAGTAAAG 436

Qy 235 TTGCTTGAAGACAGAGAGAGGACATATGAGATTAACCAAGCTTTCGAGCG 294

Dh 437 TTGCTTGAAGACAGAGAGAGGACATATGAGATTAACCAAGCTTTCGAGCG 496

Qy 295 GCAAGGCAACCCCTTTGGGTCCCTCTTTGATGAGAGCTGTTTCAATGCAAT 354

Dh 497 GCAAGGCTGCCCCCTTTGTGTCCCTCTTTGATGAGAGCTGTTTCAATGCAAT 546

Qy 355 TCACTATTAAATCTTGAACAGCACTTTCGATCATGTTTCTTAAGCTTGAAGCTG 414

|||||
Db 547 TCACCTATTAATCTCACTGACCTCTCTGTGTGTTGTTGAGTCTG 606
Qy 415 AGCTTGGCTGGCATTCACACAGCTGTTGGCCGACCGAGACCGCGCTGACTCC 474
Db 607 AGCTTGGCTGGCGCTCCACACAGCTGTTGGCCGACCGAGACCTGCGCTGACTTC 666
Qy 475 CATCCCTCTGATCATGACGAGGTGTCC 501
Db 667 CATCCCTCGGATCTGCGACGAGGTGTTC 693
RESULT 10
AAS65964
ID AAS65964 standard; cDNA; 893 BP.
XX
AC AAS65964;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #1768.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN M0200175067-A2.
XX
PD 11-OCT-2001.
XX
PP 30-MAR-2001; 2001MO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR P-PSDB; A8G01777.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 1768; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridization probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 893 BP; 223 A; 264 C; 202 G; 204 T; 0 other;
XX
XX

Query Match 75.4%; Score 377.8; DB 23; Length 893;
Best Local Similarity 88.6%; Pred. No. 4,7e-119;
Matches 451; Conservative 0; Mismatches 37; Indels 21; Gaps 3;
Qy 4 TCGGCCAACCTCCCCACACAGCACTTAGTTTCTCTGTAGATGGGGAGCTGAGACAG 63
Db 188 TCGGCCAACCTCCCCACACAGCACTTAGTTTCTCTGTAGATGGGGAGCTGAGACAG 247
Qy 64 GACTAGCTGATTTCTTAGGCTGATAAGATCCCTAAGCTAGCTGGAGAGTACAC 123
Db 248 GACTAGCTGATTTCTTAGGCTGATAAGATCCCTAAGCTAGCTGGAGAGTACAC 307
Qy 124 ATCCACCTTTAACAAGGGGCTTGCACCTTAGCTCAACCTGACCATC-----AG 174
Db 308 ATCCACCTTTAACAAGGGGCTTGCACCTTAGCTCAACCTGACCATCAGTAAGAAAG 367
Qy 175 AGAGCTACCTAAATGCTAATTAGCAAAAGCAGGAGGTAAAGAAATAGCCATCATCTA 234
Db 368 AGAGCTACCTAAATGCTAATTAGCAAAAGCAGGAGGTAAAGAAATAGCCATCATCTA 427
Qy 235 TTGCTTGAGACACAGAGAGGACATGATCGGATATTAACCAAGTCTTGAGCCG 294
Db 428 TTGCTTGAGACACAGAGGAGGACATGATGAGATATTAACCAAGGCTTGAGCCG 487
Qy 295 GCAACGGCAACCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTGTGTTTCATGCTATT 354
Db 488 GCAACGACTACCTCTTTGGGTCCCTTCCCTTTGTATGGAGCTGTGTTTCATGCTATT 537
Qy 355 TCACCTATTAAATCTTGCAACATG--CACTCTGAGTCCATGTTTCTTAGGCTTGAGC 412
Db 538 TCACCTATTAAATCTTGCAACATGCTTTGCTGTGTGTGTTGTTAGGCTTGAGC 597
Qy 413 TGAGCTTTGCTGCGCATTCACCACTGCTGTTTGCAGCAGGACCCGCGCTGACT 472
Db 598 TGAGCTTTGCTGCGCATTCACCACTGCTGTTTGCAGCAGGACCCGCGCTGACT 657
Qy 473 CCATCCCTCTGATCATGACGAGGTGTCC 501
Db 658 TCACCCCTCCAGATATGCGAGGAGGTGTCC 686
RESULT 11
AAS55630
ID AAS55630 standard; DNA; 2781 BP.
XX
AC AAS55630;
XX
DT 29-MAY-2001 (first entry)
XX
XX Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
XX Envelope protein; HERV; syncytia formation; placental development;
XX syncytia; cancer; cell adhesion; ss.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 762..2378
FT /*tag= a
FT /product= "envelope protein"
PN M0200116171-A1.
PD 08-MAR-2001.
PP 01-SEP-2000; 2000MO-FR02429.
XX
XX 01-SEP-1999; 99FR-0011141.
PR 15-SEP-1999; 99FR-0011793.
XX
XX (INRM) BIO MERIEUX
XX (INRM) INST NAT SANTE & RECH MEDICALE.
XX

XX Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
PI
XX WPI: 2001-22676/23.
DR P-PSDB; AAB67652.
XX
PT Detecting expression of human endogenous retrovirus envelope protein in
PT cells of a tissue or culture, from its ability to induce syncytia -
XX
PS Disclosure; Page 44-45; 57pp; French.
XX
CC The present sequence encodes a human endogenous retrovirus envelope
CC protein. The specification describes a method for detecting expression
CC of an envelope protein from a human endogenous retrovirus (HERV), in
CC cells, of a tissue or culture. The method comprises detecting syncytia
CC formation due to the fusogenic properties of the envelope protein.
CC Envelope polypeptides and polynucleotides are used to produce
CC therapeutic or prophylactic compositions, particularly for treatment of
CC cancer, to correct defects in placental development (or other natural
CC formation of other types of syncytia), and to promote adhesion of cells
CC in grafts or cellular repair processes. Expression of sequences
CC antisense to the polynucleotide are used to prevent formation of
CC syncytia.
XX
SQ Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 other;
Query Match 75.2%; Score 377; DB 22; Length 2781;
Best Local Similarity 100.0%; Pred. No. 1.6e-118; Indels 0; Gaps 0;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 TCGGCAACCTCCCAACAGCAGCCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACG 63
Db 2387 TCGGCAACCTCCCAACAGCAGCCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACG 2446
QY 64 GACTAGCTGATTTCTCTAGGCTGATAGATCCCTAAGCCTAGTGGAGGTGACAC 123
Db 2447 GACTAGCTGATTTCTCTAGGCTGATAGATCCCTAAGCCTAGTGGAGGTGACAC 2506
QY 124 ATCCACCTTTAAACAGCGGGGCTTGCACTAGCTCAGACCTGACCAATCAGAGCTCAC 183
Db 2507 ATCCACCTTTAAACAGCGGGGCTTGCACTAGCTCAGACCTGACCAATCAGAGCTCAC 2566
QY 184 TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 243
Db 2567 TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 2626
QY 244 AGCAGCAGAGGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCGCA 303
Db 2627 AGCAGCAGAGGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCGCA 2686
QY 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGTGTTTCAATGCTATTCTAT 363
Db 2687 ACCCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGTGTTTCAATGCTATTCTAT 2746
QY 364 TAAATCTTGCAACTGCA 380
Db 2747 TAAATCTTGCAACTGCA 2763
RESULT 12
AAK77526
ID AAK77526 standard; cDNA; 2946 BP.
XX
XX AAK77526;
XX
XX 10-AUG-1999 (first entry)
XX
XX Human secreted protein AJ172_2 cDNA.
XX
XX Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
XX bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
XX cell proliferation; cell differentiation; suppressor; tumour inhibition;
XX haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;

KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
XX Homo sapiens.
XX
XX W09926972-A1.
XX
XX 03-JUN-1999.
XX
XX 17-NOV-1998; 98WO-US24614.
XX
XX 20-OCT-1998; 98US-0175928.
XX
XX 21-NOV-1997; 97US-0976110.
XX
XX 18-MAY-1998; 98US-0080478.
XX
XX (GEMV) GENETICS INST INC.
XX
XX Collins-Racie LA, Evans C, Jacobs K, Lavallie ER,
PI McCoy JM, Werberg D, Treacy M;
XX
XX WPI: 1999-357813/30.
XX
XX P-PSDB; AA08622.
XX
XX New polynucleotides encoding secreted proteins
XX
XX Claim 13a; Page 100-101; 142pp; English.
XX
XX This invention describes novel human secreted proteins encoded by
XX polynucleotides isolated from human adult testes, adult brain, adult
XX blood or adult placenta, or murine adult bone marrow or thymus cDNA
XX libraries. The products of the invention are predicted to have biological
XX activities which would make them suitable for treating, preventing or
XX ameliorating medical conditions in humans and animals, although no
XX supporting data is given. Suggested activities include nutritional
XX activity, cytokine and cell proliferation/differentiation activity,
XX immune stimulating (e.g. as vaccines) or suppressing activity,
XX haematopoiesis regulating activity, tissue growth activity,
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX activity, cadherin/tumour invasion suppressor activity, and tumour
XX inhibition activity. The polynucleotides are also stated to be useful
XX for gene therapy.
SQ
Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;
Query Match 74.9%; Score 375.4; DB 20; Length 2946;
Best Local Similarity 99.7%; Pred. No. 6e-118;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 TCGGCAACCTCCCAACAGCAGCCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACG 63
Db 2553 TCGGCAACCTCCCAACAGCAGCCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACG 2612
QY 64 GACTAGCTGATTTCTCTAGGCTGATAGATCCCTAAGCCTAGCTGGAGGTGACAC 123
Db 2613 GACTAGCTGATTTCTCTAGGCTGATAGATCCCTAAGCCTAGCTGGAGGTGACAC 2672
QY 124 ATCCACCTTTAAACAGCGGGGCTTGCACTAGCTCAGACCTGACCAATCAGAGCTCAC 183
Db 2673 ATCCACCTTTAAACAGCGGGGCTTGCACTAGCTCAGACCTGACCAATCAGAGCTCAC 2732
QY 184 TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 243
Db 2733 TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 2792
QY 244 AGCAGCAGAGGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCGCA 303
Db 2793 AGCAGCAGAGGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCGCA 2852
QY 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGTGTTTCAATGCTATTCTAT 363
Db 2853 ACCCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGTGTTTCAATGCTATTCTAT 2912

Qy 364 TAAATCTTGCAACTGCA 380
Db 2913 TAAATCTTGCAACTGCA 2929

RESULT 13
AAZ59468
ID AAZ59468 standard; cDNA; 2946 BP.

XX AAZ59468;

XX 11-APR-2000 (first entry)

DE Human secreted protein Aji172_2 polynucleotide sequence.

XX Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
XX placental pathology; metastasis inhibitor; nutritional activity;
XX immune stimulator; haematopoiesis regulator; tissue growth;
XX tumour inhibitor; anti-inflammatory; clone Aji172_2; ATCC_98115;
XX gene therapy; ss.

XX Homo sapiens.

XX MO9960020-A1.

XX 25-NOV-1999.

XX 17-MAY-1999; 99WO-US10915.

XX 18-MAY-1998; 98US-0080478.

XX 20-OCT-1998; 98US-0175928.

XX (GEMV) GENETICS INST INC.

XX Jacobe K, McCoy JM, LaValle ER, Collins-Racie LA, Evans C;

XX Medbery D, Mi S, Treacy M;

XX WPI; 2000-116311/10.

XX P-PSDB; AAY67313.

XX New polynucleotides encoding secreted cDNA libraries, used to develop
XX products for the diagnosis and treatment of neoplastic disease

XX Claim 14; Page 107-108; 1499p; English.

XX This is the human secreted protein Aji172_2 nucleotide sequence, obtained
XX from a human adult testes cDNA library. The invention relates to secreted
XX human and murine proteins. The polynucleotides and proteins are predicted
XX to have biological activities which would make them suitable for
XX treating, preventing or ameliorating medical conditions in humans and
XX animals. Detection of the levels of the proteins can be used for the
XX diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
XX which modulate the expression or function of the proteins may be used for
XX treating a neoplastic disease and inhibiting metastasis. Other suggested
XX activities include nutritional activity (e.g. in feeds), cytokine and
XX cell proliferation/differentiation activity, immune stimulating
XX (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
XX activity, tissue growth activity, activin/inhibin activity,
XX chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
XX receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
XX invasion suppressor activity, and tumour inhibition activity. The
XX polynucleotide sequences are also stated to be useful for gene therapy.

XX Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;

XX Query Match 74.9%; Score 375.4; DB 21; Length 2946;

XX Best Local Similarity 99.7%; Pred. No. 66-118;

XX Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TGGGCAACCTCCCAAGCAGCACTTGGCTTCCTTGAATGGGGAGCTAGAGACAG 63
Db 2553 TGGGCAACCTCCCAAGCAGCACTTGGCTTCCTTGAATGGGGAGCTAGAGACAG 2612

Qy 64 GACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTTAAGCTTAGCTGGGAAGTGACAC 123
Db 2613 GACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTTAAGCTTAGCTGGGAAGTGACAC 2672
Qy 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAACCTGACCAATCAGAGAGCTCAC 183
Db 2673 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAACCTGACCAATCAGAGAGCTCAC 2732
Qy 184 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCAATCATATTGCTGAG 243
Db 2733 TAAATGCTAATTAGGCAAGAAACAGAGGTAAAGAAATAGCAATCATATTGCTGAG 2792
Qy 244 AGCAGCAGCAGAGGCAATATGATCGGATATAAACCAAGCTTCGAGCCGGCAAGGCA 303
Db 2793 AGCAGCAGCAGAGGCAATATGATCGGATATAAACCAAGCTTCGAGCCGGCAAGGCA 2852
Qy 304 ACCCCCTTTGGGGTCCCTCCCTTGTATGGAGCTGTGTTTCATGCTATTCACTCAT 363
Db 2853 ACCCCCTTTGGGGTCCCTCCCTTGTATGGAGCTGTGTTTCATGCTATTCACTCAT 2912
Qy 364 TAAATCTTGCAACTGCA 380
Db 2913 TAAATCTTGCAACTGCA 2929

RESULT 14
AAD24195
ID AAD24195 standard; cDNA; 2930 BP.

XX AAD24195;

XX 07-MAY-2002 (first entry)

XX Human syncytin cDNA.

XX Human; syncytin; pre-eclampsia; gestational trophoblast disorder;

XX choriocarcinoma; hydatiform mole; placental site tumour; abortion;

XX envelope gene; human endogenous defective retrovirus; HERV-W; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 930..2546

XX FT /*tag= a

XX FT /product= "Syncytin"

XX MO200204678-A2.

XX 17-JAN-2002.

XX 09-JUL-2001; 2001WO-US21719.

XX 07-JUL-2000; 2000US-216657P.

XX (GEMV) GENETICS INST INC.

XX Keith JC, McCoy JM, Mi S;

XX WPI; 2002-117127/22.

XX P-PSDB; AAE1540.

XX Identifying a compound for treating a subject with or at risk of
XX developing pre-eclampsia, comprises determining whether the expression
XX of activity of syncytin in the cell is modulated in the presence of a
XX test compound

XX Disclosure; Page 39-42; 43p; English.

XX The invention relates to identifying compounds which are modulators
XX of syncytin expression. The syncytin modulators are useful in diagnosis
XX and treatment of pre-eclampsia and gestational trophoblast disorders (e.g.
XX choriocarcinoma, hydatiform mole, placental site tumour and missed/
XX incomplete abortion). Syncytin is a human gene derived from the

CC envelope gene of human endogenous defective retrovirus, HERV-W. The
CC present invention is based partly on the discovery that syncytin
CC expression is dramatically reduced in preclampsia, and is also
CC mis-localised to the apical syncytiotrophoblast membrane. The present
CC sequence is human syncytin cDNA.

SQ Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 other;

Query Match 74.7%; Score 374.4; DB 24; Length 2930;
Best Local Similarity 99.7%; Pred. No. 1.3e-117;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCGGCCAACCTCCCAACAGCAGCCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 2555 TCGGCCAACCTCCCAACAGCAGCCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 2614
QY 64 GACTAGCTGATTTCTTCTGAGCTGATTAAGATCCCTAAGCTAGTGGGAAGGTACAC 123
DB 2615 GACTAGCTGATTTCTTCTGAGCTGATTAAGATCCCTAAGCTAGTGGGAAGGTACAC 2674
QY 124 ATCCACCTTTAAACACGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAG 183
DB 2675 ATCCACCTTTAAACACGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAG 2734
QY 184 TAAATGCTAATTAGGCAAGAAGAGAGTAAAGAAATAGCCATCATCTATTGCTGAG 243
DB 2735 TAAATGCTAATTAGGCAAAAACAGAGAGTAAAGAAATAGCCATCATCTATTGCTGAG 2794
QY 244 AGCAGAGAGAGAGAGCAATGATCGGGATTAACCCCAAGTCTTGAGCCGGCAACGGCA 303
DB 2795 AGCAGAGAGAGAGAGCAATGATCGGGATTAACCCCAAGTCTTGAGCCGGCAACGGCA 2854
QY 304 ACCCCCTTTGGGTCCTCCCTCTTGTATGGGAGCTGTGTTTCATGCTATTCTACTAT 363
DB 2855 ACCCCCTTTGGGTCCTCCCTCTTGTATGGGAGCTGTGTTTCATGCTATTCTACTAT 2914
QY 364 TAAATCTTGCAACTGC 379
DB 2915 TAAATCTTGCAACTGC 2930

RESULT 15
AAK25665
ID AAK25665 standard; cDNA to mRNA; 7582 BP.

XX AAK25665;

XX 21-MAY-1999 (first entry)

XX Complete human endogenous retrovirus W genome.

KM Clone; human endogenous retrovirus; genome; autoimmune disease;
KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

XX Human endogenous retrovirus.

XX W09902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-FR01442.

XX 07-JUL-1997; 97FR-0008815.

XX (INMR) BIO MERIEUX.

XX Beseme F, Blond JL, Boucon O, Mallet F, Mandrand B;

XX WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -
PT expressed exclusively in placenta and useful in diagnosis and

PT therapy of autoimmune disease, and abnormal or failed pregnancy
XX
PS Claim 1; Page 71-74; 106pp; French.

XX This sequence represents the complete sequence of the human endogenous
CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC peptides encoded by them are markers of autoimmune disease (e.g. multiple
CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC insulin-dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility.

SQ Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;

Query Match 74.5%; Score 373.4; DB 20; Length 7582;
Best Local Similarity 97.6%; Pred. No. 4.9e-117;
Matches 368; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCCAACCTCCCAACAGCAGCCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 7206 TCGGCCAACCTCCCAACAGCAGCCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 7265
QY 64 GACTAGCTGATTTCTTCTGAGCTGATTAAGATCCCTAAGCTAGTGGGAAGGTACAC 123
DB 7266 GACTAGCTGATTTCTTCTGAGCTGATTAAGATCCCTAAGCTAGTGGGAAGGTACAC 7325
QY 124 ATCCACCTTTAAACACGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAG 183
DB 7326 ATCCACCTTTAAACACGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAG 7385
QY 184 TAAATGCTAATTAGGCAAGAAGAGAGTAAAGAAATAGCCATCATCTATTGCTGAG 243
DB 7386 TAAATGCTAATTAGGCAAGAAGAGAGTAAAGAAATAGCCATCATCTATTGCTGAG 7445
QY 244 AGCAGAGAGAGAGAGCAATGATCGGGATTAACCCCAAGTCTTGAGCCGGCAACGGCA 303
DB 7446 AGCAGAGAGAGAGAGCAATGATCGGGATTAACCCCAAGTCTTGAGCCGGCAACGGCA 7505
QY 304 ACCCCCTTTGGGTCCTCCCTCTTGTATGGGAGCTGTGTTTCATGCTATTCTACTAT 363
DB 7506 ACCCCCTTTGGGTCCTCCCTCTTGTATGGGAGCTGTGTTTCATGCTATTCTACTAT 7565
QY 364 TAAATCTTGCAACTGCA 380
DB 7566 TAAATCTTGCAACTGCR 7582

Search completed: April 19, 2003, 12:18:34
Job time : 205.12 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 12:03:58 ; Search time 34.6897 Seconds
(without alignments)
4429.120 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 1 gtcctgcgacaccccccacccca.....ctgcatcatgaggggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents: NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	375.4	74.9	2946	US-09-175-928-3	Sequence 3, Appli
2	238.4	47.6	279	US-08-686-878A-50	Sequence 50, Appli
3	238.4	47.6	279	US-08-721-489-4	Sequence 4, Appli
4	46.8	9.3	7218	US-08-233-463-14	Sequence 14, Appli
5	34.2	6.8	80246	US-09-078-294-4	Sequence 4, Appli
6	34.2	6.8	80595	US-09-078-294-3	Sequence 3, Appli
7	34	6.8	7218	US-08-233-463-14	Sequence 14, Appli
8	33	6.6	1859	US-08-691-563C-46	Sequence 46, Appli
9	32	6.4	1417	US-08-117-362-31	Sequence 31, Appli
10	32	6.4	1417	US-08-486-924-31	Sequence 31, Appli
11	32	6.4	1722	US-08-691-563C-58	Sequence 58, Appli
12	30.8	6.1	9763	US-08-973-273-1	Sequence 1, Appli
13	30.6	6.1	11827	US-09-733-455-3	Sequence 3, Appli
14	30.2	6.0	1043	US-09-422-576D-6	Sequence 6, Appli
15	30.2	6.0	1091	US-09-422-576D-5	Sequence 5, Appli
16	30.2	6.0	1165	US-09-422-576D-25	Sequence 25, Appli
17	30.2	6.0	1831	US-09-422-576D-1	Sequence 1, Appli
18	29.4	5.9	29598	US-09-341-587-6	Sequence 6, Appli
19	29.2	5.8	289	US-08-896-164-16	Sequence 16, Appli
20	29.2	5.8	1132	US-08-894-731-3	Sequence 3, Appli
21	29	5.8	2023	US-08-961-083-199	Sequence 199, App
22	28.8	5.7	33768	US-08-793-559-2	Sequence 2, Appli
23	28.8	5.7	1066	US-08-793-559-2	Sequence 2, Appli
24	28.8	5.7	2142	US-08-793-559-3	Sequence 3, Appli
25	28.8	5.7	3568	US-09-218-363-3	Sequence 3, Appli
26	28.8	5.7	3825	US-08-737-597-1	Sequence 1, Appli
27	28.8	5.7	3825	US-08-737-597-2	Sequence 2, Appli

28	28.6	5.7	3552	4	US-09-157-210-3	Sequence 3, Appli
29	28.4	5.7	6836	4	US-08-976-259-73	Sequence 73, Appli
30	28.2	5.6	623	4	US-09-385-982-52	Sequence 52, Appli
31	28.2	5.6	16389	4	US-09-741-154-3	Sequence 3, Appli
32	28.2	5.6	19307	4	US-08-836-022A-10	Sequence 10, Appli
33	28.2	5.6	19307	4	US-09-427-048A-10	Sequence 10, Appli
34	28	5.6	2364	2	US-08-838-219B-5	Sequence 5, Appli
35	28	5.6	2364	3	US-09-233-336A-5	Sequence 5, Appli
36	28	5.6	2364	4	US-09-233-752A-5	Sequence 5, Appli
37	28	5.6	2364	4	US-09-402-036-5	Sequence 5, Appli
38	28	5.6	2364	4	US-09-904-285-5	Sequence 5, Appli
39	28	5.6	2370	4	US-09-002-285-79	Sequence 79, Appli
40	28	5.6	2370	4	US-09-002-285-95	Sequence 95, Appli
41	28	5.6	2374	4	US-09-002-285-97	Sequence 97, Appli
42	28	5.6	2375	4	US-08-960-780-3	Sequence 3, Appli
43	28	5.6	2375	4	US-08-960-780-7	Sequence 7, Appli
44	28	5.6	2375	4	US-09-073-898-3	Sequence 3, Appli
45	28	5.6	2375	4	US-09-073-898-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-175-928-3
Sequence 3, Application US/09175928A
Patent No. 6312921
GENERAL INFORMATION:
APPLICANT: Jacoby, Kenneth
APPLICANT: McCoy, John W.
APPLICANT: LaValle, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Metberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Mi, Sha
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
FILE REFERENCE: 6006B.AJ172A
CURRENT APPLICATION NUMBER: US/09/175, 928A
CURRENT FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 2946
TYPE: DNA
ORGANISM: Homo sapiens
US-09-175-928-3
Query Match 74.9%; Score 375.4; DB 4; Length 2946;
Best Local Similarity 99.7%; Pred. No. 5.1e-118;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 TCGGCAACCTCCCAACAGCAGCTTAGTTCTCTGTGAGATGGGAGCTGAGACAG 63
DB 2553 TCGGCAACCTCCCAACAGCAGCTTAGTTCTCTGTGAGATGGGAGCTGAGACAG 2612
QY 64 GACTRAGCTGATTTCCCTGAGCTGATTAAGATCCCTAGCCTAGCTGGAGGTGACAC 123
DB 2613 GACTRAGCTGATTTCCCTGAGCTGATTAAGATCCCTAGCCTAGCTGGAGGTGACAC 2612
QY 124 ATTCACCTTTAAACAGGGGCTTCACTAGCTACACCTGACCAATCAGAGAGCTCAC 183
DB 2673 ATTCACCTTTAAACAGGGGCTTCACTAGCTACACCTGACCAATCAGAGAGCTCAC 2732
QY 184 TAAATGCTAATTAGGCAAGAGAGGATTAAGAAATAGCAATCATCTATTGCTGAG 243
DB 2723 TAAATGCTAATTAGGCAAGAGAGGATTAAGAAATAGCAATCATCTATTGCTGAG 2792
QY 244 AGCAGAGAGAGAGGAGCAATGATCGGATATAAACCCCAAGTCTTGAGCCGGCAACGGCA 303
DB 2793 AGCAGAGAGAGAGGAGCAATGATCGGATATAAACCCCAAGTCTTGAGCCGGCAACGGCA 2852

QY 304 ACCCCCTTGGGTCCTCCCTTGTATGAGAGCTTGTTCATGCTATTTCACCTAT 363
DB 2853 ACCCCCTTGGGTCCTCCCTTGTATGAGAGCTTGTTCATGCTATTTCACCTAT 2912
QY 364 TAAATCTTGCACTGCA 380
DB 2913 TAAATCTTGCACTGCA 2929

RESULT 2

US-08-686-878A-50
Sequence 50, Application US/08686878A
Patent No. 5708157
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallee, Edward
APPLICANT: Racie, Lisa
APPLICANT: Werberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Evans, Cheryl
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/686,878A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-686-878A-50

Query Match 47.6%; Score 238.4; DB 1; Length 279;
Best Local Similarity 92.4%; Pred. No. 8.2e-72;
Matches 242; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

QY 119 ACCACATCCACTTTAAACAGGGGCTTGCACTTACACCTGACCAATCAGAG 178
DB 1 RCACATCCACTTTAAACAGGGGCTTGCACTTACACCTGACCAATCAGAG 60
QY 179 CTCACATAATGCTATTGCGCAAGACGAGGTAAATAATGCCATCTATTGC 238
DB 61 NTCANTAAATGATATTTGGCAAAAAGAGGTAAATAATAGCAATCTATTGC 120
QY 239 CTGAGAGCAGACGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCGGCAA 298
DB 121 CTGAGAGCAGACGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCGGCAA 180
QY 299 CGGCAACCCCTTTGGGTCCCTCCTTGTATGGAGTCTGTTCATGCTATTTCAC 358

DB 181 CGGCAACCCCTTTGGGTCCCTCCTTGTATGGAGCTTGTTCATGCTATTTCAN 240
QY 359 TCTATTAATCTTGCACTGCA 380
DB 241 TTTATTAATNTTGCACCTGCA 262

RESULT 3

US-08-721-489-4
Sequence 4, Application US/08721489
Patent No. 5786465
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallee, Edward
APPLICANT: Racie, Lisa
APPLICANT: Werberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721,489
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-721-489-4

Query Match 47.6%; Score 238.4; DB 1; Length 279;
Best Local Similarity 92.4%; Pred. No. 8.2e-72;
Matches 242; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

QY 119 ACCACATCCACTTTAAACAGGGGCTTGCACTTACACCTGACCAATCAGAG 178
DB 1 RCACATCCACTTTAAACAGGGGCTTGCACTTACACCTGACCAATCAGAG 60
QY 179 CTCACATAATGCTATTGCGCAAGACGAGGTAAATAATGCCATCTATTGC 238
DB 61 NTCANTAAATGATATTTGGCAAAAAGAGGTAAATAATAGCAATCTATTGC 120
QY 239 CTGAGAGCAGACGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCGGCAA 298
DB 121 CTGAGAGCAGACGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCGGCAA 180
QY 299 CGGCAACCCCTTTGGGTCCCTCCTTGTATGGAGCTGTTCATGCTATTTCAN 358
DB 181 CGGCAACCCCTTTGGGTCCCTCCTTGTATGGAGCTGTTCATGCTATTTCAN 240


```
RESULT 7
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEFFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-F18
; US-08-232-463-14

Query Match      6.8%; Score 34; DB 1; Length 7218;
Best Local Similarity 7.1%; Pred. No. 0.3;
Matches 22; Conservative 155; Mismatches 135; Indels 0; Gaps 0;
```

```
Qy 342 TTTTCATGCTAT 353
Db 1032 AATTAATCTGT 1021

RESULT 8
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glaucia PARANOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTI
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-691-563C-46

Query Match      6.6%; Score 33; DB 3; Length 1859;
Best Local Similarity 69.2%; Pred. No. 0.32;
Matches 45; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
```

```
Qy 14 TCCCAACAGACTTAGTTCTTCTGTTGAGATGGGGACTGAGACGAGCTAGCTGG 73
Db 1687 TTCCCAACAGACGCTGGGCTGTCCGTTTGAAGTGGGATTGAAGCGTGAAGCCAGCTGG 1746

Qy 74 ATTTC 78
Db 1747 ACTTC 1751

RESULT 9
US-08-117-362-31/c
; Sequence 31, Application US/08117362
; Patent No. 5595872
; GENERAL INFORMATION:
; APPLICANT: Metzerau II, John R.
; APPLICANT: Sharp, Daru Y.
; APPLICANT: Gregg, Richard E.
```


TITLE OF INVENTION: MICROSMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/117,362
FILING DATE: 03-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 783..890
US-08-117-362-31

Query Match 6.4%; Score 32; DB 1; Length 1417;
Best Local Similarity 45.6%; Pred. No. 0.6;
Matches 154; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 71 TGAATTCCTAGGCTGACTAAGATCCCTAAGCTGAGGAGTGACCAATCCACC 130
Db 1345 TGCCCTTTTCAATGGGCTAATCTCTTGAGGTCAATGCTAAGCACTAAAGAAA 1286

QY 131 TTTAAACCGGGGCTTGCACTAGCTACACCTGACCAATCAGAGAGCTACTAAATG 190
Db 1285 TTTAAGAACTGGGTTTAAAAAAATCTCATTTACAAATGAAGAGCTGCATT 1226

QY 191 CTAATTAGGCAAGACAGAGGTAAGAAATAGCCATC---ATCTATTGCTGAGAGC 246
Db 1225 CAATGTGACCAACAGAGGTAACATAGTGGTATCTTTTCTCTTTTAAAAAC 1166

QY 247 ACAGCAGAGGAGCAATGATCGGATATATAACCAAGTCTTGAGCCGCAACGGCAAC 306
Db 1165 AGACCATGAATGAGAAATTTGAAGCAAGACACTGTCTGAGAACTGAAATTAAGA 1106

QY 307 CCCCTTTGGGTCCTCCCTTTGTATGGAGCTCTTTTCAATGCTATTTCACCTATTA 366
Db 1105 AATATCTGTTTCTTAATGTTTGAAGGATTTGAATTTGTTTCATTTCACCTAGCA 1046

QY 367 ATCTTGCACTGCACTCTTGTGTCATGTTTCTTACG 404
Db 1045 TGCTTTCACAAACATTCCAAGTTACATGTTTGAAG 1008

RESULT 10
US-08-486-924-31/C
; Sequence 31, Application US/08486924
; Patent No. 5789197
; GENERAL INFORMATION:
; APPLICANT: Wetterau II, John R.
; APPLICANT: Sharp, Daru Y.
;

APPLICANT: Gregg, Richard E.
TITLE OF INVENTION: MICROSMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117,362
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 783..890
US-08-486-924-31

Query Match 6.4%; Score 32; DB 1; Length 1417;
Best Local Similarity 45.6%; Pred. No. 0.6;
Matches 154; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 71 TGAATTCCTAGGCTGACTAAGATCCCTAAGCTGAGGAGTGACCAATCCACC 130
Db 1345 TGCCCTTTTCAATGGGCTAATCTCTTGAGGTCAATGCTAAGCACTAAAGAAA 1286

QY 131 TTTAAACCGGGGCTTGCACTAGCTACACCTGACCAATCAGAGAGCTACTAAATG 190
Db 1285 TTTAAGAACTGGGTTTAAAAAAATCTCATTTACAAATGAAGAGCTGCATT 1226

QY 191 CTAATTAGGCAAGACAGAGGTAAGAAATAGCCATC---ATCTATTGCTGAGAGC 246
Db 1225 CAATGTGACCAACAGAGGTAACATAGTGGTATCTTTTCTCTTTTAAAAAC 1166

QY 247 ACAGCAGAGGAGCAATGATCGGATATATAACCAAGTCTTGAGCCGCAACGGCAAC 306
Db 1165 AGACCATGAATGAGAAATTTGAAGCAAGACACTGTCTGAGAACTGAAATTAAGA 1106

QY 307 CCCCTTTGGGTCCTCCCTTTGTATGGAGCTCTTTTCAATGCTATTTCACCTATTA 366
Db 1105 AATATCTGTTTCTTAATGTTTGAAGGATTTGAATTTGTTTCATTTCACCTAGCA 1046

QY 367 ATCTTGCACTGCACTCTTGTGTCATGTTTCTTACG 404
Db 1045 TGCTTTCACAAACATTCCAAGTTACATGTTTGAAG 1008

RESULT 11
US-08-691-563C-58
; Sequence 58, Application US/08691563C
;

```

; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-RADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-691-563C-58

Query Match          6.4%; Score 32; DB 3; Length 1722;
Best Local Similarity 73.2%; Pred. No. 0.66;
Matches 41; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 GTCTGGCCCAACCTCCCAACAGCACTTAGTTCCTGTTGAGATGGGGACTGA 56
Db 1024 GTCATCGGCCAATTCCCAACAGCAGTTGGGTGCTCTGTTTAGAGGGGGGATGA 1079

RESULT 12
US-08-973-273-1/c
; Sequence 1, Application US/08973273
; Patent No. 6140085
; GENERAL INFORMATION:
; APPLICANT: Dean, Caroline
; APPLICANT: MacKnight, Richard C
; APPLICANT: Bancroft, Ian
; APPLICANT: Lister, Clare K
; TITLE OF INVENTION: Genetic Control of Flowering
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderyhe P.C.
; STREET: 1100 No. 6140085th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,273
; FILING DATE: 01-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01332
; FILING DATE: 03-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 951196.9
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ms Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9763 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEITICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; STRAIN: Columbia
; US-08-973-273-1

Query Match          6.1%; Score 30.8; DB 3; Length 9763;
Best Local Similarity 63.5%; Pred. No. 4.4;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 162 CCTGACCAATCAGAGAGCTCTAAATGCTAATTAGCAAGACAGAGGTAAGAAT 221
Db 4845 CCTTCAGAAAGAAAGAACTATATCATCTGATAGTAAAGCAAGAAACAGGAAAGAAA 4786

Qy 222 AGCCAATCATCTAT 235
Db 4785 GGATCAACATCCAT 4772

RESULT 13
US-09-739-455-3
; Sequence 3, Application US/09739455
; Patent No. 6413756
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000653
; CURRENT APPLICATION NUMBER: US/09/739,455
; CURRENT FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 11827
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(11827)
; OTHER INFORMATION: n = A,T,C or G
; US-09-739-455-3

Query Match          6.1%; Score 30.6; DB 4; Length 11827;
Best Local Similarity 51.1%; Pred. No. 5.8;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
```

Qy	103	CTTACGTGGGAAGGAGACCAATCCACCTTTAAACAACGGGGCTTGCAACTTAGTCACAC	162
Db	10853	CCACCTGTGTGAGAGTCGGAACGCTGCGCTTTTGTGGCGGCTTGTTCTTAAATCAGTT	10912
Qy	163	CTGACCAATCAGAGAGCTCACTAAATGCTAATTAGGCANAAGACAGAGGTAAAGATA	222
Db	10913	CCCTCTTAGGATTTTATTAACATCAAAAAAAAAAATTAAGTTTGTAAAAGAAATGAGAGATA	10972
Qy	223	GCCAATCATCTATTGCGCTGAG	243
Db	10973	CAGAAACATGAATTTACAGAG	10993

RESULT 14
US-09-422-576D-6

```

: Patent No. 6395549
:
: GENERAL INFORMATION:
: APPLICANT: Tuan, Dorothy
: APPLICANT: Long, Qiaoming
: APPLICANT: Bengira, Chikh
: TITLE OF INVENTION: Long Terminal Repeat, Enhancer, and Insulator Sequences for Use
: TITLE OF INVENTION: Recombinant Vectors
: FILE REFERENCE: M0351-205010
: CURRENT APPLICATION NUMBER: US/09/422,576D
: CURRENT FILING DATE: 1999-10-21
: PRIOR APPLICATION NUMBER: US 60/105,256
: PRIOR FILING DATE: 1998-10-22
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 6
: LENGTH: 1043
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-422-576D-6

```

	Query Match	Best Local Similarity	Score 6.0%;	Pred. 62.1%;	DB 30.2;	Length 1043;
	Matches	64; Conservative	0;	No.Matches	38; Indels	1; Gaps
Qy	253	GGAGGAGCAATGATCGGAGATATATAACCCAAAGCTTTGAGCGCGGCAACGGCAACCCCTTT	312			
Db	485	GGTGGGCGCCAGATTAAGAGATATAAGACAGGCTGCCGCGACCAAGCACTGGCAACCGGC-TC	543			
Qy	313	GGGTCCTCCCTCTTTGTATGGGAGCTCTGTTTCAATGATATT	355			
Db	544	GGGTCCCTTCCACTGTAAGCTTTATCTTTGGCTCTTT	586			

```

RESULT 15
US-09-422-576D-5
: Sequence 5, Application US/09422576D
: Patent No. 6395549
: GENERAL INFORMATION:
: APPLICANT: Tuan, Dorothy
: APPLICANT: Long, Qiaoming
: APPLICANT: Bengira, Chikh
: TITLE OF INVENTION: Long Terminal Repeat, Enhancer, and Insulator Sequences for Use
: TITLE OF INVENTION: Recombinant Vectors
: FILE REFERENCE: M0351-205010
: CURRENT APPLICATION NUMBER: US/09/422,576D
: CURRENT FILING DATE: 1999-10-21
: PRIOR APPLICATION NUMBER: US 60/105,256
: PRIOR FILING DATE: 1998-10-22
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 5
: LENGTH: 1091
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus Sequence

```

```
; NAME/KEY: misc_feature
; LOCATION: (437)..(437)
; OTHER INFORMATION: n = any nucleotide
US-09-422-576D-5
```

Query Match	Similarity	6.0%	Score 30.2	DB 4	Length 1091
Best Local	Similarity 62.1%	Pred. No. 2.1			
Matches 64	Conservative 0	Mismatches 38	Indels 1	Gaps 1	
Qy	253	GGAGGACAATGATGGGATATAAACCAAGCTTCGAGCGGCAACGCCCTTT	312		
Db	534	GGTGGGGCCAGTAAAGAGATATAAGCAGGCTGCCCGAGCCAGCAGTGGCAACCCGC	-TC 592		
Qy	313	GGGTCCTCTCTTGTATGGAGGCTGTGTTTCATGCTATT	355		
Db	553	GGGTCCCTTCCACACTGTGGAGCTTTGTCTTCCTGCTTTT	635		

Search completed: April 19, 2003, 14:49:59
Job time : 94.6897 secs

```

; NAME/KEY: misc feature
; LOCATION: (437)..(437)
; OTHER INFORMATION: n = any nucleotide
US-09-422-576D-5

Query Match          6.0%; Score 30.2; DB 4; Length 1091;
Best Local Similarity 62.1%; Pred. No. 2.1;
Matches 64; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 253 GGAGGGACATGATCGGGATATTAACCCAGTCTTCGAGCCGGCAACGGCAACCCCTTT 312
    |||||
DB 534 GGTGGGGCCAGATAAGAGATTAAGCAGCCTGCCCGAGCCAGCAGTGGCAACCCGC-TC 592
    |||||

QY 313 GGGTCCCTCCCTTTGTATGGAGCTGTGTTTCATGCTATTT 355
    |||||
DB 593 GGGTCCCTCCCACTGTGGAGGCTTGTCTTTCGCTCTTT 635
    |||||

Search completed: April 19, 2003, 14:49:59
Job time : 94.6897 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 12:08:18 ; Search time 75.717 Seconds

(without alignments)
6646.925 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501
Sequence: 1 gtcgcgcacactcccca.....ctgatcatcgaggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	426	85.0	1329	7 US-08-979-847-108	Sequence 108, App
2	375.4	74.9	2946	9 US-10-114-893-134	Sequence 134, App
3	374.4	74.7	2930	10 US-09-902-535-1	Sequence 1, Appl
4	341.8	68.2	635	7 US-08-979-847-102	Sequence 102, App
5	317.6	63.4	180557	12 US-10-003-806-6	Sequence 6, Appl
6	317.6	63.4	180557	12 US-10-003-806-9	Sequence 58, Appl
7	311.4	62.2	410	10 US-09-880-107-538	Sequence 3, Appl
8	305.2	60.9	326014	10 US-09-731-231A-3	Sequence 8173, App
9	286.6	57.2	541	10 US-09-864-761-8173	Sequence 4444, App
10	281.2	56.1	1894	10 US-09-864-761-4444	Sequence 1654, App
11	266.8	53.3	15425	9 US-10-091-504-1654	Sequence 1654, App
12	266.8	53.3	15425	9 US-09-764-869-1654	Sequence 50, Appl
13	238.4	47.6	279	12 US-10-040-916-50	Sequence 14951, A
14	203.8	40.7	569	10 US-09-864-761-14951	Sequence 20462, A
15	146.2	29.2	246	10 US-09-864-761-20462	Sequence 3, Appl
16	136.8	27.3	446	10 US-09-811-284-3	Sequence 7027, Appl
17	136.2	27.2	504	10 US-09-864-761-7027	Sequence 1, Appl
18	129.6	25.9	3524	10 US-09-972-724-1	Sequence 895, App
19	124.8	24.9	475	10 US-09-864-761-895	

C 20	122.4	24.4	559	10 US-09-864-761-7501	Sequence 7501, App
C 21	111.6	22.3	440	10 US-09-864-761-3694	Sequence 3694, App
C 22	105.4	21.0	409	10 US-09-864-761-4153	Sequence 4153, App
C 23	92.2	18.4	579	10 US-09-864-761-13678	Sequence 13678, A
C 24	58.8	11.7	387	10 US-09-864-761-30194	Sequence 30194, A
C 25	58	11.6	525	10 US-09-893-737-31	Sequence 31, Appl
C 26	55.4	11.1	88	10 US-09-864-761-20907	Sequence 20907, A
C 27	50.6	10.1	275	10 US-09-864-761-17675	Sequence 17675, A
C 28	45.4	9.1	464	10 US-09-867-701-2947	Sequence 2947, App
C 29	42.6	8.5	929	9 US-10-093-766-14	Sequence 14, Appl
C 30	42.2	8.4	362	10 US-09-867-701-5411	Sequence 5411, App
C 31	41.8	8.3	1529	9 US-10-125-237-75	Sequence 75, Appl
C 32	41.2	8.2	81	10 US-09-864-761-24915	Sequence 24915, A
C 33	39.2	7.8	635	9 US-09-796-692-5087	Sequence 5087, App
C 34	39	7.8	3824	9 US-10-036-041-22	Sequence 22, Appl
C 35	39	7.8	3824	9 US-10-028-072-541	Sequence 541, App
C 36	39	7.8	3824	9 US-10-035-855-22	Sequence 22, Appl
C 37	39	7.8	3824	9 US-10-121-049-541	Sequence 541, App
C 38	39	7.8	3824	9 US-10-123-904-541	Sequence 541, App
C 39	39	7.8	3824	9 US-10-140-470-541	Sequence 541, App
C 40	39	7.8	3824	9 US-09-931-836-22	Sequence 22, Appl
C 41	39	7.8	3824	9 US-10-175-746-541	Sequence 541, App
C 42	39	7.8	3824	9 US-10-176-918-541	Sequence 541, App
C 43	39	7.8	3824	9 US-10-176-921-541	Sequence 541, App
C 44	39	7.8	3824	9 US-10-227-884-209	Sequence 209, App
C 45	39	7.8	3824	9 US-10-036-214-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-979-847-108
Sequence 108, Application US/08979847
Publication NO. US20030039664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACDEL, GLAUCIA
APPLICANT: KOMRIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUXE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979, 847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 108:

SEQUENCE CHARACTERISTICS:
LENGTH: 1329 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-108

Query Match 85.0%; Score 426; DB 7; Length 1329;
Best Local Similarity 90.6%; Pred. No. 8.2e-136;
Matches 453; Conservative 0; Mismatch 47; Indels 0; Gaps 0;

QY 2 TCTGGCCAACTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGTGAAGAC 61
DB TGTGAGCAACTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGTGAAGAC 556
QY 62 AGGATTAAGTGAATTTCTAGAGTGAATGATCTTAAGCTTACCTGGAAGTGAAC 121
DB AGGATTAAGTGAATTTCTAGAGTGAATGATCTTAAGCTTACCTGGAAGTGAAC 616
QY 122 ACATCACTTTAAACAGGGGCTTGAACCTTAAGCTCAACCTGACCAATCAGAGCTC 181
DB GCATCACTTTAAACAGGGGCTTGAACCTTAAGCTCAACCTGACCAATCAGAGCTC 676
QY 182 ACTAAATGCTAATTAGGCAAGAGAGGATTAAGGATTAAGGATTAAGGATTAAG 241
DB ACTAAATGCTAATTAGGCAAGAGAGGATTAAGGATTAAGGATTAAGGATTAAG 736
QY 242 AGAGCAAGAGAGAGAGAGATGATGAGATTAAGGATTAAGGATTAAGGATTAAG 301
DB AGAGCAAGAGAGAGAGAGATGATGAGATTAAGGATTAAGGATTAAGGATTAAG 796
QY 302 CAACCCCTTTGGGATCCCTCCCTTTGATGGAGCTGTTTTCATGTTTACTCT 361
DB CAACCCCTTTGGGATCCCTCCCTTTGATGGAGCTGTTTTCATGTTTACTCT 856
QY 362 ATTAATCTTGAACCTGACTCTTCTGTCATGTTCTTAAGGCTTGAAGCTTTC 421
DB ATTAATCTTGAACCTGACTCTTCTGTCATGTTCTTAAGGCTTGAAGCTTTC 916
QY 422 GCTGCCATCCCACTGCTGTTTGGCCGACCGCAACCCGCTGACTCCATCCCT 481
DB GCTGCCATCCCACTGCTGTTTGGCCGACCGCAACCCGCTGACTCCATCCCT 976
QY 482 CTGATCATGAGGGTGTCC 501
DB CTGATCATGAGGGTGTCC 996

RESULT 2

US-10-114-893-134
Sequence 134, Application US/10114893
Publication No. US20020193567A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallee, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Bowman, Michael R.
APPLICANT: Spaulding, Vikki
APPLICANT: Carlin-Duckett, Mckeough
APPLICANT: Kelleher, Kerry S.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYPEPTIDES ENCODING THEM
FILE REFERENCE: GI 6000-10A
CURRENT APPLICATION NUMBER: US/10/114,893
EARLIER FILING DATE: 2002-04-02
CURRENT FILING DATE: 1999-10-06
NUMBER OF SEQ ID NOS: 321

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 134
LENGTH: 2946
TYPE: DNA
ORGANISM: Homo sapiens
US-10-114-893-134

Query Match 74.9%; Score 375.4; DB 9; Length 2946;
Best Local Similarity 99.7%; Pred. No. 3e-118;
Matches 376; Conservative 0; Mismatch 1; Indels 0; Gaps 0;

QY 4 TCGGCCAACTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGTGAAGAC 63
DB TCGGCCAACTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGTGAAGAC 2612
QY 64 GACTAGCTGATTTCTAGAGTGAATGATCTTAAGCTTACCTGGAAGTGAAC 123
DB GACTAGCTGATTTCTAGAGTGAATGATCTTAAGCTTACCTGGAAGTGAAC 2672
QY 124 ATCACTTTAAACAGGGGCTTGAACCTTAAGCTCAACCTGACCAATCAGAGCTC 183
DB ATCACTTTAAACAGGGGCTTGAACCTTAAGCTCAACCTGACCAATCAGAGCTC 2732
QY 184 TAAATGCTAATTAGGCAAGAGAGGATTAAGGATTAAGGATTAAGGATTAAG 243
DB TAAATGCTAATTAGGCAAGAGAGGATTAAGGATTAAGGATTAAGGATTAAG 2792
QY 244 AGCAGAGAGAGAGAGATGAGGATTAAGGATTAAGGATTAAGGATTAAG 303
DB AGCAGAGAGAGAGAGATGAGGATTAAGGATTAAGGATTAAGGATTAAG 2852
QY 304 ACCCCCTTTGGGATCCCTCCCTTTGATGGAGCTGTTTTCATGTTTACTCT 363
DB ACCCCCTTTGGGATCCCTCCCTTTGATGGAGCTGTTTTCATGTTTACTCT 2912
QY 364 TAAATCTTGAACCTGCA 380
DB TAAATCTTGAACCTGCA 2929

RESULT 3

US-09-902-535-1
Sequence 1, Application US/09902535
Patent No. US20020102530A1
GENERAL INFORMATION:
APPLICANT: Keith, Jr., James C.
APPLICANT: McCoy, John M.
TITLE OF INVENTION: Methods and compositions for diagnosing
TITLE OF INVENTION: and treating preclampsia and gestational trophoblast
FILE REFERENCE: GIN-6006B4
CURRENT APPLICATION NUMBER: US/09/902,535
CURRENT FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: 60/216,657
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2930
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (930)...(2546)
US-09-902-535-1

Query Match 74.7%; Score 374.4; DB 10; Length 2930;
Best Local Similarity 99.7%; Pred. No. 6.6e-118;
Matches 375; Conservative 0; Mismatch 1; Indels 0; Gaps 0;

QY 4 TCGGCCAACTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGTGAAGAC 63
|||||

Db 2555 TCGGCAACTCCCAACAGCATTAGCTTTCTGTGAGATGGGGACCTGAGAGACAG 2614
Qy 64 GACTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGGAAGTACAC 123
Db 2615 GACTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGGAAGTACAC 2674
Qy 124 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCACAACCTGACCAATCAGAGACTCAG 183
Db 2675 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCACAACCTGACCAATCAGAGACTCAG 2734
Qy 184 TAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTTAG 243
Db 2735 TAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTTAG 2794
Qy 244 AGCAGCAGAGAGGAGCAATGATCGGATATATTAACCAAGTCTTGAGCCGGCAACGGCA 303
Db 2795 AGCAGCAGAGAGGAGCAATGATCGGATATATTAACCAAGTCTTGAGCCGGCAACGGCA 2854
Qy 304 ACCCCCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTCTTTCATGCTATTTCCTCTAT 363
Db 2855 ACCCCCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTCTTTCATGCTATTTCCTCTAT 2914
Qy 364 TAAATCTTGCAACTGC 379
Db 2915 TAAATCTTGCAACTGC 2930

RESULT 4
US-08-979-847-102
; Sequence 102, Application US/08979847
; Publication No. US200300396441
; GENERAL INFORMATION:
; APPLICANT: BERON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUXE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-979-847-102
Query Match 68.2%; Score 341.8; DB 7; Length 635;
Best Local Similarity 94.2%; Pred. No. 4.6e-107;
Matches 355; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 4 TCGGCAACTCCCAACAGCATTAGCTTTCTGTGAGATGGGGACCTGAGAGACAG 63
Db 243 TCAGCAACTCCCAACAGCATTAGCTTTCTGTGAGATGGGGACCTGAGAGACAG 302
Qy 64 GACTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGGAAGTACAC 123
Db 303 GACTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGGAAGTACAC 362
Qy 124 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCACAACCTGACCAATCAGAGACTCAG 183
Db 363 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCACAACCTGACCAATCAGAGACTCAG 422
Qy 184 TAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTTAG 243
Db 423 TAAATGCTAATTAGGCAAAATAGAGGTAAAGAAATAGCCAAATCATCTATTGCTTAG 482
Qy 244 AGCAGCAGAGAGGAGCAATGATCGGATATATTAACCAAGTCTTGAGCCGGCAACGGCA 303
Db 483 AGCAGCAGAGAGGAGCAATGATCGGATATATTAACCAAGTCTTGAGCCGGCAACGGCA 542
Qy 304 ACCCCCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTCTTTCATGCTATTTCCTCTAT 363
Db 543 ACCCCCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTCTTTCATGCTATTTCCTCTAT 602
Qy 364 TAAATCTTGCAACTGCA 380
Db 603 TAAATCTTGCAACTGAA 619

RESULT 5
US-10-003-806-6/c
; Sequence 6, Application US/10003806
; Patent No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulnik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02066US1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-6
Query Match 63.4%; Score 317.6; DB 12; Length 180557;
Best Local Similarity 85.9%; Pred. No. 1.9e-97;
Matches 395; Conservative 0; Mismatches 44; Indels 21; Gaps 3;
Qy 52 ACTGAGACACGAGCTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGG 111
Db 58991 AGTGAAGACAGGACTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGG 58932
Qy 112 GAAGGTACACATCACTTAAACACGGGGCTTGCACTTAGCTCACAACCTGACCAAT 171
Db 58931 GAAGGTACACCTTCCACTTTAAACATGGGGCTTGCACTTAGCTCACAACCTGACCAAT 58972
Qy 172 C-----AGAGGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATA 222

[illegible]

```

RESULT 6
US-10-003-806-9/c
; Sequence 9, Application US/10003806
; Patent No. US2002011929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulnik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P020606US/10024824
; CURRENT APPLICATION NUMBER: US/10/003.806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: SnapGene version 3.1
; SEQ ID NO 9
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-9

```

Query Match	63.4%	Score 317.6;	DB 12;	Length 180557;
Best Local Similarity	85.9%	Pred. No. 1.5e-97;		
Matches 395;	Conservative 0;	Mismatches 44;	Indels 21;	Gaps 3.

QY	52	AAGGAGAGCAGAGCTAGCTGATTTTCCATAGGCTAGCTAAGAAATCCCTAAGCTGAGCTGG	111
Db	58991	AGTGAAGAGCAGAGACTAGCTGAATTTCTTAGGCCGAGCTAAGAAATCCCTAAGCTGAGCTGG	58932
QY	112	GAAAGTGACCAATCCACCTTTAAACACGGGGCTTGCACCTTAGCTCACACCTGCACCAAT	171
Db	58931	GAAAGTGACCGCTTCCACCTTTAAACATGGGGCTTGCACCTTAGCTCACACCGGACCAAT	58872
QY	172	C-----AGAGAGCTACTAAATGCTAATTAGGAAAGACAGAGAGTAAAGAAATA	222
Db	58871	CAGATAGTAAAGAGAGCTACTAAATAGTTAATTAGGAAAGAAACAGAGAGTAAAGAAATA	58812
QY	223	GCCAAATCATCTAATTGCTGAGAGCACAGAGAGAGGAAATATGGGGATATAAACCCAA	282
Db	58811	GCCAAATCATCTAATTGCTGAGAGCACAGAGAGAGGGAATATGATCAGCATATAAACCCAG	58752
QY	283	GTCTTCGAGCCGGCAACGGCAACCCCTTTGGGTCCCCCTCTTGTATAGGAGCTCTGT	342
Db	58751	GCATCTGAGCCGAGCAACAGACTAGGCTCTTTGGGTCCCCCTCTTGTATAGGAGCTCTGT	58692
QY	343	TTTCAATGCATTTTCACTCTATTAATCTTGAACATGCA--CTCTCTGGTTCGATGTTTCT	400
Db	58691	-----CTTCACTCTAATTAAATCTTGAACATGCACTCTCTTTTGGTCTCAATTTGT	58642

Oy 401 TAGCGGTAGAGGACTTTCGCTGCGCATCCACCATGCTGTTTGGCGCACCGGAGAC 460
 Db 58641 CATGTTTGAGCTGAGACTTTCCTGCGCGTCCACCATGCTGTTTGGCGCTGCGGAGAC 58881
 Oy 461 CCGCGCGCTGACTCCCATCCCTCTGAGATCATGAGGATGTC 500
 Db 58581 CTGCTGCTGACTTCATCCGTCGTCGCAATCCGGACAGATGTC 58542

```

RESULT 7
US-09-880-107-538/c
/ Sequence 538. Application US/09880107
/ Patent No. US20020142981A1
/ GENERAL INFORMATION:
/ APPLICANT: Horne, Darci T.
/ APPLICANT: Vockley, Joseph G.
/ APPLICANT: Scherf, Uwe
/ APPLICANT: Gene Logic, Inc.
/ TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
/ FILE REFERENCE: 44921-5028-WO
/ CURRENT APPLICATION NUMBER: US/09/880,107
/ CURRENT FILING DATE: 2001-06-14
/ PRIOR APPLICATION NUMBER: US 60/211,379
/ PRIOR FILING DATE: 2000-06-14
/ PRIOR APPLICATION NUMBER: US 60/237,054
/ PRIOR FILING DATE: 2000-10-02
/ NUMBER OF SEQ ID NOS: 3950
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 538
/ LENGTH: 410
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA250958
US-09-880-107-538

```

Query Match	62.2%	Score 311.4;	DB 10;	Length 410;
Best Local Similarity	92.6%;	Pred. No. 1e-95;		
Matches 349;	Conservative 0;	Mismatches 26;	Indels 2;	Gaps 2;

QY	4	TCGGCAACCTCCCCAACAGACACTAGATTTTCTGTGAGATGGGGGACTGACAGACAG	63
Db	377	TCAGCAACTCCACACAGACACTTGGGTTTTCTGTGAGAGGGGACTGAGAGACAG	318
QY	64	GACTAGCTGATTTCTTAGGCTGACTAAGATCCCTAAGCTAGCTGGGAAGTGACAC	123
Db	317	GACTAGCTGATTTCTTAGGCGATTAAGATCCCTAACCCTAGCTGGGAAGTGACCGC	256
QY	124	ATCCACTTTTAAACAGGGGCTTGCAATTAGCTCACACTGACCAATCAGAGGCTCAC	183
Db	257	ATCCACTTTTAAACAGGGGCTTGCAACTTAGCTCACACCACCAATCAGAGGCTCAC	198
QY	184	TAAATGCTAATTGAGCAACAGAGGTAAAGAAATGACATCATATTGCTCGA-	242
Db	197	TAAATGCTAATTGAGCAAAACAGAGGTAAAGAAATAGCCAATCATATTGCTCGAG	133
QY	243	GAGCAGACAGAGAGGACAATGATCGGGATATAACCAAGTCTCGAGCGCGC-AAACGG	301
Db	137	GAGCAGAGTGGGAGGAGCAAGGATTTGCATTTAAACCAGGCATTCGAGCCAGGAAACGG	78
QY	302	CAACCCCTTTGGGTCCTCCCTTGTATGGAGCTCTGTGTTCAATGTAATTTCACTCT	361
Db	77	CAACGCTTTGGGTCCTCCCTTGTATGGAGCTCTGTGTTCACTAATTTCACTCT	18
QY	362	ATTAAATCTTGCAACTG	378
Db	17	ATTAAATCTTGCAACTG	1

RESULT 8
US-09-731-231A-3/c
; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1


```
;; GENERAL INFORMATION:
;; APPLICANT: GUEGLER, Karl et al
;; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;; TITLE OF INVENTION: THEREOF
;; FILE REFERENCE: CL001007
;; CURRENT APPLICATION NUMBER: US/09/731,231A
;; CURRENT FILING DATE: 2000-12-07
;; NUMBER OF SEQ ID NOS: 6
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 326014
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1) - (326014)
;; OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3
```

```
Query Match          60.9%; Score 305.2; DB 10; Length 326014;
Best Local Similarity 89.1%; Pred. No. 4.7e-93;
Matches 344; Conservative 0; Mismatches 33; Indels 9; Gaps 1;
```

```
QY 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTTGAATGGGGAGTGAAGACAG 63
Db 170580 TCGGCACTCTCCCAACAGCACTTAGGTTTCTGTTGAATGGGGAGTGAAGACAG 170521

QY 64 GACTAGCTGATTTCTCTAGGCTGATAGATTCCTAAGCCTAGCTGGAAAGTACCAAC 123
Db 170520 GACTAGCTGATTTCTCTAGGCTGATAGATTCCTAAGCCTAGCTGGAAAGTACCTGC 170461

QY 124 ATCCACCTTTAAACAGCGGGGGCTTGCACTTAGCTCAGACCTGACCAATC-----AG 174
Db 170460 TTCTACCTTTAAACCGGGGGCTTGCACTTAGCTCAGACCTGACCAATGAGTGAAG 170401

QY 175 AGAGCTCAATAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCCATCATCTA 234
Db 170400 AGAGCTCAATAATGCTAATTTAGGCTAAGCAAGAGGTAAAGAAATAGCCATCATCTA 170341

QY 235 TTGCTGAGACACAGCAGAGGAGCAATGATCGGATATTAACCCAGTCTTGAGCCG 294
Db 170340 TTGCTGAGACACAGCAGAGGAGCAATGATCGGATATTAACCCAGCATTGAGACCA 170281

QY 295 GCAATGCGCAACCCCTTTGGGTCCTCCCTTGTATGGGAGCTGTGTTTCATGCTAAT 354
Db 170280 GCAATGCGTACCATTTTGGGTCCTCCCTTGTATGGGAGCTGTGTTTCATGCTAAT 170221

QY 355 TCACCTATTAATCTTCAACTGCA 380
Db 170220 TCACCTATTAATCTTCAACTGCA 170195
```

```
RESULT 9
US-09-864-761-8173/c
; Sequence 8173, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
```

```
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 8173
;; LENGTH: 541
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC016663.2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
US-09-864-761-8173
```

```
Query Match          57.2%; Score 286.6; DB 10; Length 541;
Best Local Similarity 82.7%; Pred. No. 3.9e-88;
Matches 383; Conservative 0; Mismatches 59; Indels 21; Gaps 4;
```

```
QY 49 GGAAGTGAAGACAGGACTAGCTGATTTCTTAGGCTGATTAAGATTCCTAAGCTTAGC 108
Db 536 GGCAGTGAAGACAGGACTAGCTGATTTCTTAGGCTGATTAAGATTCCTAAGCTTAGC 477

QY 109 TGGGAAGGTGACCAATCCACTTTAAACAGGGGCTTGGCACTTAGCTCACACTGACC 168
Db 476 TGGGAAGGTGACCAATCCACTTTAAACAGGGGCTTGGCACTTAGCTCACACTGACC 417

QY 169 AATC-----AGAGAGCTCAATAAATGCTAATTTAGGCAAGACAGAGGTAAAGA 219
Db 416 AATCAGCTAGTAAAGACAGCTCAATAAAGCTAATTCGGCTAATAACAGAGGTAAATA 357

QY 220 ATAGCAATCATCTAATGCTGAGAGCAGCAGAGGAGCAATGATCGGATATTAACC 279
Db 356 ATAGCAATCATCTAATGCTGAGAGCAGCAGAGGAGTAAATGATCGGATATTAACC 297

QY 280 CAAGCTTTCAGCGCGGAGCAAGCAACCCCTTGGGTCCTCCCTTGTATGGAGGCTC 339
Db 296 CAGGATTCAGCGCGGAGCAATCAAGCGCTTAGGGTCCTCCCACTGTATGGAGGCTC 237

QY 340 TGTTCATGCTAATTTACTCTAATTAATTTGCAACTGCACTCT-TCTGATCATGTTT 398
```

Db 236 TGT-----TTTCACTATTAATCTTG-2ACTGCACACTCTGTGCTGTGTTT 187
Qy 399 CTTAGCGTTAGCTGACCTTTCGTCGCCATCAACCACTGTTTCCGCCACCGAG 458
Db 186 ATTCCGTTCAATGACCTTTCTCCACCAATCAACCACTGTAATGCTGTGAG 127
Qy 459 AC-CGCGCTGACTCCGATCCCTCTGATCATCAGGGTGTG 500
Db 126 ACACCGCTGACTTCCACCCCTCCGATCTGCAGGGTGTG 84
RESULT 10
US-09-864-761-4444
Sequence 4444, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 4444
LENGTH: 1894
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC002346.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELM, SIGNAL = 7.2
US-09-864-761-4444
Query Match 56.1%; Score 281.2; DB 10; Length 1894;
Best Local Similarity 88.4%; Pred. No. 5,4e-86;
Matches 320; Conservative 0; Mismatches 33; Indels 9; Gaps 1;
Qy 2 TCTGCGCCAACTTCCCAACAGCACTTAGTTTCTGTTGAGATGGGGCACTAGAGAC 61
Db 1457 TGTGGCCAACTTCCCAACAGCACTTAGTTTCTGTTGAGAGGGGGCACTAGAGAC 1516
Qy 62 AGGACTAGCTGATTTCTTCTAGCTGATAGATCCCTTAGCTGAGGAGTGACC 121
Db 1517 AGGATATCTAGATTTCTTAGCACTTAAGATCCCTTAGCTGAGGAGTGACC 1576
Qy 122 ACATCCACCTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAATC----- 172
Db 1577 GCTTCCACCTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAATCTAA 1636
Qy 173 AGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATACCAATCATC 232
Db 1637 AGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATACCAATCATC 1696
Qy 233 TATTGCTGAGAGCAAGCAGAGAGGACAAATGATCGGATTTAAACCAATCTTCGAG 292
Db 1697 TATTGCTGAGAGCAAGCAGAGAGGACAAATGATCGGATTTAAACCAATCTTCGAG 1756
Qy 293 CGGCAGCGCAACCCCTTTGGTCCCTCCCTTTGATGGAAGCTGTGTTGATGCTA 352
Db 1757 CAGTACAGCTACCTCTTTGGTCCCTCCCTTTGATGGAAGCTGTGTTGATGCTA 1816
Qy 353 TT 354
Db 1817 TT 1818
RESULT 11
US-10-091-504-1654/c
Sequence 1654, Application US/10091504
Publication No. US2003005908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1654
LENGTH: 15425
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1654
Query Match 53.3%; Score 266.8; DB 9; Length 15425;
Best Local Similarity 88.1%; Pred. No. 1,4e-80;
Matches 317; Conservative 0; Mismatches 32; Indels 11; Gaps 2;
Qy 4 TCGGCCAACTTCCCAACAGCACTTAGTTTCTGTTGAGATGGGGCACTAGAGAG 63
Db 2049 TTGGCCAACTTCCCAACAGCACTTAGTTTCTGTTGAGAGGGGGAGCT--GAGACAG 1992
Qy 64 GACTAGCTGATTTCTTCTAGCTGATAGATCCCTTAGCTGAGGAGGATGACAC 123
Db 1991 AACTAGCTGATTTCTTCTAGCTGATAGATCCCTTAGCTGAGGAGGATGACAC 1932

Qy	124	ATCACCTTTAAACACGGGGCTTGCACTTAGCTCACCTGACCAATC-----AG	174
Db	1931	ATCCATCTTTAAACATGGGGCTTGCACTTAGCTCACACCCGACCAATAGGATTAAG	1872
Qy	175	AGAGCTCACTAAAATGCTAAATTAGGCAAAACAGAGGTAAAGAAATAGCCATCATCTA	234
Db	1871	AGAGCTCACAGAAATGCTAATTAGGCAAAAACAGAGGTAAACAAATAGCCATCATCTA	1812
Qy	235	TTTGCTTAGAGCACAGAGGAGGACATATATGGGATTTAAACCCAAAGCTTTGACGGC	294
Db	1811	TCGGCTTAGAGCACAGGAGGAGACATATATGGGATTTAAACCCGAGCATTTAGGTG	1752
Qy	295	GCAACGCAACCCCTTTGGGATCCCTCCCTTTGTATGAGGAGCTGTGTTTCATGCTATT	354
Db	1751	GCAACAGTACCTCTTTGGATCCCTCCCTTTGTATAGAGCTGTGTTTACATCTACT	1692

```

RESULT 12
US-09-764-869-1654/c
: Sequence 1654, Application US/09764869
: Patent No. US2002006151A1
: GENERAL INFORMATION:
: APPLICANT: ROSEN ET AL.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC007
: CURRENT APPLICATION NUMBER: US/09/764,869
: PRIORITY FILING DATE: 2001-01-17
: Prior application data removed - refer to PALM or file wrapper
: NUMBER OF SEQ ID NOS: 2442
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1654
: LENGTH: 15425
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-869-1654

```

Query Match	Similarity	53.3%	Score 266.8	DB 10	Length 15425
Best Local	Similarity	88.1%	Pred. Nc. 1.4e-80		
Matches	317	Conservative	0	Mismatches 32	Indels 11
					Gaps 2
Qy	4	TCGGCCAACTTCCCAACAGCACTTAGTGTCTTGTGAGATGGGGGACTGAGACAG	63		
Db	2049	TTGGCCAACTTCCCAACAGCACTTGGGTTTCTGTGAGGGGGGACT--GAGACAG	1992		
Qy	64	GACTAGCTGGATTTCTTAGGCTTGACTAAGAAATCCCTAAGCTTGGAGGTGCAC	123		
Db	1991	AACTAGCTGGATTTCTTAGCTAGTCCAGCTAAGAAATGCTTAAGCTTGGAGGTGCAC	1932		
Qy	124	ATCCACCTTTTAAACAGGGGGCTTGACAATTACTCACACCTGACCAATC-----AG	174		
Db	1931	ATCCACCTTTTAAACAGGGGGCTTGACAATTACTCACACCTGACCAATAGGTTATTAAG	1872		
Qy	175	AGAGCTCACTTAAATAGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCATCTA	234		
Db	1871	AGAGCTCACAGAAATCTTAATTAGGCAAAAGACAGAGGTAAACAAATAGCCATCATCTA	1812		
Qy	235	TTGCTTGAGAGCACAGACGAGGAGCAATGATCGGGATTTAAACCCAAATCTTTGACCG	294		
Db	1811	TCGCTTGAGAGCACAGTGGAGGAGCAATGATCGGGATTTAAACCCAGGCAATTGAGCTG	1752		
Qy	295	GGAAGCGGACCCCTTTGGGTCCCTCCCTTTGTATGGAGACTCTGTTTCAATGCTATT	354		
Db	1751	GGAAGCGGACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATT	1692		

RESULT 13
US-10-040-916-50
; Sequence 50, Application US/10040916
; Patent NO. US20020146769A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John
; Lavallie, Edward
;

```

1  Racie, Lisa
2  Werberg, David
3  Treacy, Maurice
4  Evans, Cheryl
5  Spaulding, Vikki
6  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
7  ENCODING THEM
8  NUMBER OF SEQUENCES: 71
9  CORRESPONDENCE ADDRESS:
10 ADDRESS: Genetics Institute, Inc.
11 STREET: 87 Cambridgepark Drive
12 CITY: Cambridge
13 STATE: Massachusetts
14 COUNTRY: U.S.A.
15 ZIP: 02140
16
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: Patent Invention #1.0, Version #1.30
22
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/10/040,916
25 FILING DATE: 07-Jan-2002
26 CLASSIFICATION: <Unknown>
27
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: 08/887,029
30 FILING DATE: 07-FEB-1997
31 APPLICATION NUMBER: 08/686,878
32 FILING DATE: <Unknown>
33
34 ATTORNEY/AGENT INFORMATION:
35 NAME: Brown, Scott A.
36 REGISTRATION NUMBER: 32,724
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: (617) 498-8224
39 TELEFAX: (617) 876-5851
40
41 INFORMATION FOR SEQ ID NO: 50:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH: 279 base pairs
44 TYPE: nucleic acid
45 STRANDEDNESS: double
46 TOPOLOGY: linear
47
48 MOLECULE TYPE: cDNA
49
50 SEQUENCE DESCRIPTION: SEQ ID NO: 50:
51
52 US-10-040-916-50

```

Query Match	Similarity	47.6%	Score 238.4	DB 12	Length 279
Best Local	Similarity	92.4%	Pred. No. 9.9e-72		
Matches	Conservative	1	Mismatches 19	Indels	Gaps
Qy	119	ACCAATCCACCTTTAAACACGGGGCTTGCACCTAGCTCACCTGCACCATCAGAG	178		
Db	1	RCACATCCACCTTTAAACACGGGGNTTGCAGAAAAGATMCACTTGACCATCAGAG	60		
Qy	179	CTCACTAAATGCTAATTATGAGCAAGAGGTTAAAGAAATAGCCCATCATCTATTGC	238		
Db	61	NTCANTAAATGATATATTATTTGGCAAAAAACAGAGGTTAAAGAAATAGCCCATCATCTATTGC	120		
Qy	239	CTGAGAGCACACGACGAGAGGACATGATCGGATTTAAACCCCAAGTCTTGCAGCCGGCAA	298		
Db	121	CTGAGAGCACACGACGAGAGGACATGATCGGATTTAAACCCCAAGTCTTNGAGCCGGCAA	180		
Qy	239	CGGCAACCCCTTTGGGTGGCCCTCCCTTTGATATGGGACCTCTGTTTCATGCTATTTCAC	358		
Db	181	CGGCAACCCCTTTGGGTGGCCCTCCCTTTGATATGGGAGCTGTTTCATGCTATTTCAN	240		
Qy	359	TCTATTAAATCTTGCAACTGCA	380		
Db	241	TNTATTAAATNTTGCAACTGCA	262		

RESULT 14
US-09-864-761-14951/C
; Sequence 14951, Application US/09864761

Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT FILING DATE: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14951
LENGTH: 569
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP00233.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER. SIGNAL = 9.3
OTHER INFORMATION: EXPRESSED IN PLACENTA. SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW. SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER. SIGNAL = 2.1
US-09-864-761-14951
Query Match 40.7%; Score 203.8; DB 10; Length 569;
Best Local Similarity 79.3%; Pred. No. 1.1e-59;
Matches 337; Conservative 0; Mismatches 62; Indels 26; Gaps 7;

Qy 171 TC-----AGAGACTCACTAAATGCTAATTAGGC-AAACAGAGAGGTAAAGAA 220
Db 295 TCAGTAGTAAAGAGGGTTCACCTAATAATTAAGCTTAAAGAGAGGTAAAGAA 236
Qy 221 TAG-CCATCATCTATTGCTGAGAGCAGAGAGAGGAGCAATGATCGGATATTAAC 279
Db 235 TAGTAAATCATATATGCTTAAACAGCAGAGAGGAGGAGCAATGATTTGGATTTAAAC 176
Qy 280 CAA--GTCTTGAGCCGCGCAACGCCATCCCTTTGGTCCCTCCCTTTGTATGGAGC 337
Db 175 CAGGAGATTCAACCGGAGAGGCGCAACCCCTTTGGGTCCCTCCCATTTATGGAGC 116
Qy 338 TCTGTTTCATGCTATTTCATCTATTAAATCTTGCACTG--CACTTTCTGCTCATG 395
Db 115 TCTGT-----TTTCACTCTGTAAATCTTGCAACTGTAACTCTCTGTGCTAGTG 66
Qy 396 TTTCTTACGCTTGAGCTGAGCTTTCGCTGCGATCCACCACTGCTGTTTCCGCCACG 455
Db 65 TTTGTTCCGCTCAGAGCTGAACCTTTTGCTCACCTCTTACCACTGCTGTTCCCGGTAG 6
Qy 456 CAGAC 460
Db 5 CAGAC 1

RESULT 15
US-09-864-761-20462/c
Sequence 20462, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT FILING DATE: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30

```
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20462
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010951.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
; OTHER INFORMATION: NT HIT: AB026898.1, EVALU6 4.00e-57
; OTHER INFORMATION: EST HUMAN HIT: AI492055.1, EVALU6 3.00e-49
; OTHER INFORMATION: SWISSPROT HIT: Q02279, EVALU6 7.30e-01
US-09-864-761-20462

Query Match      29.2%; Score 146.2; DB 10; Length 246;
Best Local Similarity 88.5%; Pred. No. 4.4e-40;
Matches 185; Conservative 0; Mismatches 13; Indels 11; Gaps 2;

Qy 232 CTATTGCTGAGAGCAGCAGGAGGAGGACATGATCGGGATATTAACCCAGTCTTCGAG 291
Db 246 CTGTTGCTGAGAGCAGCAGGAGGAGGACATTAATCAGGATATAAACCCAGCATTCGAG 187
Qy 232 CCGGCAAGGCAACCCCTTGGGTCCCTCCCTTGTATGGAGCTCTGTTTCATGCT 351
Db 186 CTGGCAACGTAACCCCTTGGGTCCCTCCCTTGTATGGAGCTCTAT----- 136
Qy 352 ATTCACTTATTAAATCTTGCACTGCACTCTTCTGGTCCATGTTCTTAACGGCTTGAG 411
Db 135 -CTTCACTTATTAAATCTTGCACTGCACTCTTCTGGTCCGTTGTTAC-GCTTGAG 78
Qy 412 CTGAGCTTTGCTGCGCATCCACCACTGC 440
Db 77 CTGAGCTTTGCTGCGCATCCACCACTGC 49
```

Search completed: April 19, 2003, 14:58:25
Job time : 257.717 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 11:45:23 ; Search time 1268.51 Seconds
(without alignments)
6396.429 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501
Sequence: 1 gctcgcgcacaccccca.....ctggatcatgcaggggtgcc 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	376	75.0	440	12	BE732673
2	372.8	74.4	494	9	AA781423
3	372.2	74.3	436	9	AI128526
4	365.8	73.0	415	9	AI128496
5	360.4	71.9	388	14	H01325
6	355.4	70.9	727	17	AG029908

C	7	348.4	69.5	422	14	NS3177	NS3177 yv56h11.s1
	8	345.2	68.9	771	13	BI087886	BI087886 602852690
	9	343.8	68.6	683	17	AG102951	AG102951 Pan trogl
C	10	339.6	67.8	385	14	NS5091	NS5091 yv43e03.s1
	11	337.8	67.4	458	14	R76086	R76086 y171b03.s1
C	12	332.4	66.3	522	10	AW971553	AW971553 EST383642
	13	329.6	65.7	363	14	T69704	T69704 yd13a03.s1
C	14	326.6	65.2	342	9	AA860368	AA860368 aj59c05.s
C	15	326.4	65.1	424	14	R27412	R27412 yb46d11.s1
C	16	324.4	64.8	342	14	T47345	T47345 yb10h02.s1
C	17	318.8	63.6	438	14	R72728	R72728 y175d06.s1
C	18	317.6	63.4	443	9	AA837267	AA837267 od26b10.s
C	19	312.6	62.4	433	9	AI379210	AI379210 tcd01g11.x
C	20	312.6	62.2	490	9	AI598135	AI598135 tnl4a10.x
C	21	311.4	62.2	410	9	AA250958	AA250958 z807d10.s
C	22	311	62.1	425	9	AI570707	AI570707 tm79g09.x
C	23	309.4	61.8	446	9	AI393478	AI393478 tga5g04.x
C	24	307.2	61.3	921	17	BH189565	BH189565 ENTQ48TR
C	25	305.4	61.0	463	14	R68685	R68685 y14g06.s1
C	26	302.6	60.4	404	14	R27389	R27389 yb46a09.s1
C	27	296.2	59.1	328	9	AA729556	AA729556 nx58c05.s
C	28	293.6	58.6	485	10	AW511366	AW511366 h445h03.x
C	29	293	58.5	431	9	AA552941	AA552941 nk61a10.s
C	30	290.8	58.0	471	9	AA709471	AA709471 zt91h06.s
C	31	289.6	57.8	653	17	AG033781	AG033781 Pan trogl
C	32	287.8	57.4	701	17	AG126669	AG126669 Pan trogl
C	33	285	56.9	611	17	AG381711	AG381711 RPII11-16
C	34	284	56.7	712	17	AO892947	AO892947 HS 3131.B
C	35	281.2	56.1	722	17	AG049481	AG049481 Pan trogl
	36	279.2	55.7	503	14	BM723292	BM723292 UI-E-BJ0-
C	37	279.2	55.7	530	14	BM682297	BM682297 UI-E-BJ0-
C	38	278.4	55.6	470	9	AI074704	AI074704 ox83d05.s
C	39	277.6	55.4	446	9	AI288235	AI288235 q18a08.x
C	40	276.2	55.1	960	17	AO900343	AO900343 HS 3179.B
C	41	273.2	54.5	893	14	BQ437925	BQ437925 AGENCOURT
C	42	272.8	54.5	609	17	AG066901	AG066901 Pan trogl
C	43	272.4	54.4	326	14	D29167	D29167 HUMNR203 Hu
	44	272.4	54.4	447	12	BF919416	BF919416 QVO-NT015
	45	271.8	54.3	681	10	AV722664	AV722664 AV722664

ALIGNMENTS

RESULT 1
BE732673
LOCUS BE732673 440 bp mRNA linear EST 15-SEP-2000
DEFINITION 601571305F1 NIH_MGC_21 Homo sapiens CDNA clone IMAGE:3925728 5',
mRNA sequence.
ACCESSION BE732673
VERSION BE732673.1 GI:10146665
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 440)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strauberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LICW752 row: j column: 01
High quality sequence stop: 440.
Location/Qualifiers
1. .440
FEATURES
source
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone="IMAGE:3925728"
/clone_1id="N1H_MGC_21"
/l1sue_type="cHorizCar:inoma"
/lab_host="DH10B (phage resistant)"
/note="Organ: placenta; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 124 a 117 c 101 g 58 t

ORIGIN

Query Match 75.0%; Score 376; DB 12; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.8e-111;
Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTTGTGAGATGGGGAGCTGAGAGACAG 63
DB 65 TCGGCAACCTCCCAACAGCACTTAGGTTTCTTGTGAGATGGGGAGCTGAGAGACAG 124

QY 64 GACTAGCTGATTTCTTCTAGCTGACTAGATCCCTAGCTAGCTGGAGGTGACAC 123
DB 125 GACTAGCTGATTTCTTCTAGCTGACTAGATCCCTAGCTAGCTGGAGGTGACAC 184

QY 124 ATCCACCTTTAAACAGCGGGCTTGGCACTTAGCTC/CACCTGACCAATAGAGAGCTCAC 183
DB 185 ATCCACCTTTAAACAGCGGGCTTGGCACTTAGCTC/CACCTGACCAATAGAGAGCTCAC 244

QY 184 TAAATGCTAATTAGGCAAGAGAGGTAAGATATGCAATCATATTGCTGAG 243
DB 245 TAAATGCTAATTAGGCAAGAGAGGTAAGATATGCAATCATATTGCTGAG 304

QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAACTTGGAGCGGCAACGCA 303
DB 305 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAACTTGGAGCGGCAACGCA 364

QY 304 ACCCCCTTTGGGTCCTTCTTGTATGGAGCTCTGTTTTCATGCTATTTCACCTCAT 363
DB 365 ACCCCCTTTGGGTCCTTCTTGTATGGAGCTCTGTTTTCATGCTATTTCACCTCAT 424

QY 364 TAAATCTTGCAACTGC 379
DB 425 TAAATCTTGCAACTGC 440

RESULT 2
AA781423 494 bp mRNA linear EST 31-DEC-1998
LOCUS aJ2603.s1 Soares_testis_NHT Homo sapiens cDNA clone 1391428 3'
DEFINITION similar to contains PTR7.tl PRT7 repetitive element; mRNA
sequence.
ACCESSION AA781423
VERSION AA781423.1 GI:2840754
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D. Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnl.gov/bbrp/image/image.html
Insert Length: 1645 Std Error: 0.00
Seq primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 475.
Location/Qualifiers
1. 494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="1391428"
/clone_1id="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer 15'
TGTTCACCAATCTGAGAGGAGCGCGCCCAATTTTCTTTTCTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 111 c 133 g 141 t 1 others

ORIGIN

Query Match 74.4%; Score 372.8; DB 9; Length 494;
Best Local Similarity 99.2%; Pred. No. 2.1e-110;
Matches 374; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTTGTGAGATGGGGAGCTGAGAGACAG 63
DB 382 TCGGCAACCTCCCAACAGCACTTAGGTTTCTTGTGAGATGGGGAGCTGAGAGACAG 323

QY 64 GACTAGCTGATTTCTTCTAGCTGACTAGATCCCTAGCTAGCTGGAGAGGTGACAC 123
DB 322 GACTAGCTGATTTCTTCTAGCTGACTAGATCCCTAGCTAGCTGGAGAGGTGACAC 263

QY 124 ATCCACCTTTAAACAGCGGGCTTGGCACTTAGCTC/CACCTGACCAATAGAGAGCTCAC 183
DB 262 ATCCACCTTTAAACAGCGGGCTTGGCACTTAGCTC/CACCTGACCAATAGAGAGCTCAC 203

QY 262 TAAATGCTAATTAGGCAAGAGAGGTAAGATATGCAATCATATTGCTGAG 243
DB 202 TAAATGCTAATTAGGCAAGAGAGGTAAGATATGCAATCATATTGCTGAG 143

QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAACTTGGAGCGGCAACGCA 303
DB 142 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAACTTGGAGCGGCAACGCA 83

QY 304 ACCCCCTTTGGGTCCTTCTTGTATGGAGCTCTGTTTTCATGCTATTTCACCTCAT 363
DB 82 ACCCCCTTTGGGTCCTTCTTGTATGGAGCTCTGTTTTCATGCTATTTCACCTCAT 23

QY 364 TAAATCTTGCAACTGCA 380
DB 22 TAAATCTTGCAACTGCA 6

RESULT 3
A1128526 436 bp mRNA linear EST 27-OCT-1998
LOCUS g661h10.x1 Soares_Placenta_8c9weeks_2bDH81c9w Homo sapiens cDNA
DEFINITION clone IMAGE:1714147 3' similar to contains PTR7.b1 PRT5 repetitive
element; mRNA Sequence.
ACCESSION A1128526
VERSION A1128526.1 GI:3597040
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE	1 (bases 1 to 436)
NCI-CCAG	http://www.ncbi.nlm.nih.gov/ncicagp.
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE	Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: csgapbs-roman1.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lln.gov) for further information. Insert Length: 720 Std Error: 0.00 Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop: 428. Location/Qualifiers 1. 436 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1714147" /clone_lib="Soares placenta 8t09weeks 2NBHP8t09W" /dev_stage="two placenta: one from 8 weeks and another from 9 weeks post conception" /lab_host="VDH10B (ampicillin resistant)" /note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTATCCATCTGAAGTGGAGCGGCGCCGCGATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Felina Bonaldo."
FEATURES	
SOURCE	
BASE COUNT	95 a 96 c 106 g 139 t
ORIGIN	
Query Match	74.3%; Score 372.2; DB 9; Length 436;
Best Local Similarity	99.2%; Pred. No.3.1e-110;
Matches 374; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
4	TCGGCCAACTCCCAACAGCACTTAGTGTTCCTGTTGAGATGGGGGACTGAGAGACNG 63
Db	TCGGCCAACTCCCAACAGCACTTAGTGTTCCTGTTGAGATGGGGGACTGAAGAAGC 340
64	GACTGCTGAGATTTCTTAGCTGACTTAAGAAATCCCTTAAGCTTGGGAGGTACCC 123
Db	GACTGCTGAGATTTCTTAGCTGACTTAAGAAATCCCTTAAGCTTGGGAGGTACCC 280
339	GACTGCTGAGATTTCTTAGCTGACTTAAGAAATCCCTTAAGCTTGGGAGGTACCC 280
124	ATCCACCTTTAAACAGGGGGCTTGGCACTTACTACACCTTAACCAATCAGAGCTCAC 183
Db	ATCCACCTTTAAACAGGGGGCTTGGCACTTACTACACCTTAACCAATCAGAGCTCAC 220
279	ATCCACCTTTAAACAGGGGGCTTGGCACTTACTACACCTTAACCAATCAGAGCTCAC 220
184	TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTGAG 243
Db	TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTGAG 160
219	TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTGAG 160
244	AGCAGACGAGAGGGGACATGATCGGATATTAACCCAGTCTTGGAGCGGCAACGGCA 303
Db	AGCAGACGAGAGGGGACATGATCGGATATTAACCCAGTCTTGGAGCGGCAACGGCA 100
159	AGCAGACGAGAGGGGACATGATCGGATATTAACCCAGTCTTGGAGCGGCAACGGCA 100
304	ACCCCTTTGGGTCCCTCCCTTGTATGAGGAGCTGTGTTTCATGCTATTTCACCTAT 363
Db	ACCCCTTTGGGTCCCTCCCTTGTATGAGGAGCTGTGTTTCATGCTATTTCACCTAT 40
99	ACCCCTTTGGGTCCCTCCCTTGTATGAGGAGCTGTGTTTCATGCTATTTCACCTAT 40
364	TAAATCTTGCAACTGCA 380
Db	TAAATCTTGCAACTGCA 23
39	TAAATCTTGCAACTGCA 23
RESULT 4	
AI128496/c	AI128496 415 bp mRNA linear EST 27-OCT-1998
LOCUS	gc61e08.x1 Soares placenta 8t09weeks 2NBHP8t09W Homo sapiens cDNA
DEFINITION	clone IMAGE:171418 3' similar to contains PTR.11 PTR7 repetitive
ACCESSION	AI128496

KEYWORDS	VERSION
AI128496.1	GI:3597010
EST.	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS	1 (bases 1 to 415)
TITLE	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1184 Std Error: 0.00 Seq primer: -40ml3 fwd. ET from Amer sham High quality sequence stop: 413.
FEATURES	Location/Qualifiers
source	1..415
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:1714118"
	/clocation.lib="Scars placenta 8to9weeks 2NDHPeto9W"
	/dex_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
	/lab_host="DH10B (ampicillin resistant)"
	/note="Organ: placenta; Vector: pTV73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAAGTGGAGCGGCGGCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTV73 vector (Pharmacia). Library constructed by Bento Soares and M. Patricia Bonaldo."
BASE COUNT	94 a 96 c 107 g 116 t 2 others
ORIGIN	
Query Match	73.0%; Score 365.8; DB 9; Length 415;
Best Local Similarity	97.6%; Pred. No. 3.7e-108;
Matches 370; Conservative	0; Mismatches 9; Indels 0; Gaps 0;
0y	1 GTCTGGGCAACCTTCCCAAGCACTTGAAGTTTCTGTTGAAGTGGGGAAGTGAAGA 60
Db	379 GTCTGGGCAACCTTCCCAAGCACTTGAAGTTTCTGTTGAAGTGGGGAAGTGAAGA 320
0y	61 CAGAGCTAGCTGAGATTTCTTAGGCTGATTAAGATCCCTAAGCCTAGCTGGGAAGTAC 120
Db	319 CAGAGCTAGCTGAGATTTCTTAGGCTGATTAAGATCCCTAAGCCTAGCTGGGAAGTAC 260
0y	121 CACATCCACCTTTAAACAGGGGGCTTGACACTTGAAGTCTGACACTGACCAATCAGAGACT 180
Db	259 CACATCCACCTTTAAACAGGGGGCTTGACACTTGAAGTCTGACACTGACCAATCAGAGACT 200
0y	181 CACTAAATGCTAATTAGGCAAAAGCAGAGAGTAAAGAAATAGCCAAATCATCTATTGCT 240
Db	199 CACTAAATGCTAATTAGGCAAAAGCAGAGAGTAAAGAAATAGCCAAATCATCTATTGCT 140
0y	241 GAGAGCAGACAGAGGAGCAATGATCGGATATTAAACCAAGTCTTGGAGCCGCAACG 300
Db	139 GAGAGCAGACAGAGGAGCAATGATCGGATATTAAACCAAGTCTTGGAGCCGCAACG 80
0y	301 GCAACCCCTTTGGATGCCCTCCCTTTGTATGGGAGCTGTGTTTCATGCTATTGACTC 360
Db	79 GCAACCCCTTTGGATGCCCTCCCTTTGTATGGGAGCTGTGTTTCATGCTATTGACTC 20
0y	361 TATTAAATCTTGCAACTGC 379
Db	19 TATTAAATCTTGCAACTGC 1

H01325/c 388 bp mRNA linear EST 19-JUN-1995
 LOCUS Y199601.61 Soares Placenta Nb2Hp: Homo sapiens cDNA clone
 DEFINITION IMAGE:147384 3', mRNA sequence.
 ACCESSION H01325
 VERSION H01325.1
 KEYWORDS GI:864258
 SOURCE EST.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 388)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, P., Trevaaskis, E., Waterston,
 R., Williamson, A., Wohlmann, P. and Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 TITLE JOURNAL
 COMMENT
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 790
 High quality sequence scope: 346
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 790 Std Error: 0.00
 Seq primer: Promega -21ml3
 High quality sequence scope: 346.
 Location/Qualifiers
 1..388
 /organism="Homo sapiens"
 /db_xref="GDB:559031"
 /db_xref="taxon:9606"
 /clone="IMAGE:147384"
 /clone_1lb="Soares placenta Nb2Hp"
 /sex="Female"
 /dev_stage="Placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Placenta; Vector: pT73D (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 AACTGGAAGATTGGCGCGCGAGATTGTTTGTGTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 "
 BASE COUNT 90 a 90 c 96 g 110 t 2 others
 ORIGIN
 Query Match 71.9%; Score 360.4; DB 14; Length 388;
 Best Local Similarity 97.8%; Pred. No. 2.1e-106;
 Matches 364; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 244 AGCAGCAGGAGGAGCAATATGATCGGATATTAACCCAGTCTTCGACCGGCAACGGCA 303
 DB 137 AGCAGCAGGAGGAGCAATATGATCGGATATTAACCCAGTCTTCGACCGGCAACGGCA 78
 QY 304 ACCGCTTTGGGTCCTCCCTCTTTGATGAGAGCTCTGTTTCAATGCTATTTACTCTAT 363
 DB 77 AGCCCTTTGGGTCCTCCCTCTTTGATGAGAGCTCTGTTTCAATGCTATTTACTCTAT 18
 QY 364 TAAATCTTCA 375
 DB 17 TAAATCTTCA 6
 RESULT 6 727 bp DNA linear GSS 01-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-002C04.R, genomic survey sequence.
 ACCESSION AG029908
 VERSION AG029908.1
 KEYWORDS GI:16556781
 SOURCE GSS.
 ORGANISM Pan troglodytes male lymphoblast DNA, clone_1lb: PTB Chimpanzee Male
 BAC library, clone: PTB-002C04.R.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 1
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of library PTB
 Unpublished
 2 (bases 1 to 727)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (02-AUG-2001) Ageo Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 E-mail: chimpanzee@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/
 Tel: 81-45-503-9111, Fax: 81-45-503-9170
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.
 PRIMERS
 Sequencing: M13Rev
 LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 Location/Qualifiers
 1..727
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-002C04.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_1lb="PTB Chimpanzee Male BAC Library"
 "
 BASE COUNT 170 a 215 c 175 g 166 t 1 others
 ORIGIN
 Query Match 70.9%; Score 355.4; DB 17; Length 727;
 Best Local Similarity 91.1%; Pred. No. 1.2e-104;
 Matches 411; Conservative 0; Mismatches 36; Indels 4; Gaps 3;

Db 395 CGGAGAGCTCACTAAAGCTTAATAGCAAAACAGAGGTAAGAATAGCCATCAT 454
 Qy 232 CTATGCTGAGAGACAGACAGCAATGATCGGATTAATTAACCAAGCTTCGAG 291
 Db 455 CTATTGCTGAGAGACAGAGTGGAGATAGATCGGATTAATTAACCAAGCTTCGAG 514
 Qy 292 CCGGCAACGAGACCCCTTTGGTCCCTCCCTTTGATAGGAGACTGTTTCATCT 351
 Db 515 CCAGCAACGAGACCCCTTTGGTCCCTCCCTTTGATAGGAGACTGTTTCATCT 573
 Qy 352 ATTTCACCTCTTAATTAATTTGCACTGCACTCTTGGTCCATGTTTTCAGGCTTGA 411
 Db 574 ATTTCACCTCTTAATTAATTTGCACTGCACTCTTGGTCCATGTTTTCAGGCTTGA 633
 Qy 412 CTGAGCTTTCGCTGCGCAATCCACCTGCTGT-TGCCGCAACGAGACCCGCGCTGGA 470
 Db 634 CTGAGCTTTCGCTGCGCAATCCACCTGCTGT-TGCCGCAACGAGACCCGCGCTGGA 693
 Qy 471 CTCCCATCCCTCTGATCATGACAGGCTGCC 501
 Db 694 CTTCATCCCTC-GATCAGAGAGGTGTCC 722
 RESULT 7
 LOCUS N53177.1
 DEFINITION 422 bp mRNA linear EST 28-JUN-1997
 IMAGE:246789.3, similar to contains PTR1, c2 PTR7 repetitive element
 ;, mRNA sequence.
 N53177
 VERSION N53177.1 GI:1194343
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Hallier L., Lennon G., Becker M., Bonaldo M.F., Chiappelli B.,
 Chisese S., Dietrich N., Dubuque T., Favellio A., Gibb W., Hawkins
 M., Hultman M., Kucaba T., Lacy M., Le M., Le N., Mardis E., Moore
 B., Morris M., Parsons J., Prange C., Rifkin L., Rohlfing T.,
 Schelleberg K., Soares M.B., Tan F., Thierry-Mieg J., Trevisakis E.,
 Underwood K., Wohlmann P., Watson R., Wilson R. and Marra M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 TITLE
 JOURNAL
 MEDLINE
 COMMENT
 CONTACT: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 938 Std Error: 0.00
 Seq primer: ml3 -40 forward
 High quality sequence stop: 301.
 Location/Qualifiers
 1. 422
 /organism="Homo sapiens"
 /db_xref="GDB:3796035"
 /db_xref="taxon:9606"
 /clone="IMAGE:246789"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="MDH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo (dT) primer
 [5', AACTGCAAGATTAATTAAGACTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Patricia Bonaldo."
 BASE COUNT 95 a 95 c 111 g 120 t 1 others
 ORIGIN
 Query Match 69.5%; Score 348.4; DB 14; Length 422;
 Best Local Similarity 98.7%; Pred. No. 1.8e-102;
 Matches 372; Conservative 0; Mismatches 2; Indels 3; Gaps 2;
 Qy 4 TCGGCAACCTCCCAACAGACACTTATGTTTCTGTTGAGATGGGAGCTGAGACAG 63
 Db 375 TCGGCAACCTCCCAACAGACACTTATGTTTCTGTTGAGATGGGAGCTGAGACAG 316
 Qy 64 GACTAGCTGAGATTTCTAGAGCTGACTAAGATCCCTAAGCTTACTGGGAGTGACAC 123
 Db 315 GACTAGCTGAGATTTCTAGAGCTGACTAAGATCCCTAAGCTTACTGGGAGTGACAC 256
 Qy 124 ATCCACCTTTAAACAGGAGCTTCACTTACCTGACACCTGACCAATCAGAGAGCTCAC 183
 Db 255 ATCCACCTTTAAACAGGAGCTTCACTTACCTGACACCTGACCAATCAGAGAGCTCAC 196
 Qy 184 TAAATGCTAATTAGGCAAGACAGAGTAAAGAAATAGCCAAATCATTTGCTGAG 243
 Db 195 TAAATGCTAATTAGGCAAGACAGAGTAAAGAAATAGCCAAATCATTTGCTGAG 136
 Qy 244 AGCAGAGAGAGGAGCAATGATGCGATTAATTAACCAAGCTTCGAGCGGCAACGGCA 303
 Db 135 AGCAGAGAGAGGAGCAATGATGCGATTAATTAACCAAGCTTCGAGCGGCAACGGCA 77
 Qy 304 ACCGCTTTGGGCTCCCTCCCTTTGATGAGAGCTGTTTCATGCTATTTCACTCTAT 363
 Db 76 ACCGCTTTGGGCTCCCTCCCTTTGATGAGAGCTGTTTCATGCTATTTCACTCTAT 19
 Qy 364 TAAATCTTGCACTGCA 380
 Db 18 TAAATCTTGCACTGCA 2
 RESULT 8
 LOCUS B1087886
 DEFINITION 771 bp mRNA linear EST 20-JUN-2001
 IMAGE:493894.5,
 mRNA sequence.
 B1087886
 VERSION B1087886.1 GI:14506216
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 COMMENT
 Contact: Robert Strauberg, Ph.D.
 Email: gcrabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Inyte Genomics, Inc.
 DNA Sequencing by: Inyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM11015 row: d column: 23
 High quality sequence stop: 762.
 Location/Qualifiers
 1. 771
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:493894"
 /clone_lib="NIH_MGC_10"

```
/cell_line="MGC36"  
/lab_host="DH108"  
/note="Organ: Cervix; Vector: PCMV-SpOrf6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: oligo dt.  
Average insert size 1.5 kb. Library prepared by Life  
Technologies."  
BASE COUNT      225 a      206 c      166 g      174 t  
ORIGIN  
Query Match      68.9%; Score 345.2; DB 13; Length 771;  
Best Local Similarity 97.4%; Pred. No. 2.6e-101;  
Matches 372; Conservative 0; Mismatches 8; Indels 2; Gaps 2;  
QY 1 GTCCTGGCCACCTCCCAACAGCAGCATTAGGTTTCTGT-TGAGATGGGGAGCTGAGAG 59  
DB 377 CTCCTGGCCACCTCCCAACAGCAGCATTAGGTTTCTGTGAGATGGGGAGCTGAGAG 436  
QY 60 ACAGACTAGCTGATTTCTTAGGCTGACTAGAAATCCCTAAGCTTACTGGAAGTGA 119  
DB 437 ACAGACTAGCTGATTTCTTAGGCTGACTAGAAATCCCTAAGCTTACTGGAAGTGA 496  
QY 120 CCACATCCACTTTTAAACAGGGGCTTGCACTTACCTCACACCTGACCAATCAGAGAGC 179  
DB 497 CCACATCCACTTTTAAACAGGGGCTTGCACTTACCTCACACCTGACCAATCAGAGAGC 556  
QY 180 TCCTAAATGCTAATTTAGCAAGAGAGGTA-AGAATATAGCCATCATCTATTGC 238  
DB 557 TCCTAAATGCTAATTTAGCAAGAGAGGTA-AGAATATAGCCATCATCTATTGC 616  
QY 239 CTGAGAGCAGCAGAGAGGAGCAATGATCGGATTAACCCAGCTTCGACCGGCA 298  
DB 617 CTGAGAGCAGCAGAGAGGAGCAATGATCGGATTAACCCAGCTTCGACCGGCA 676  
QY 299 CGGCAACCCCTTTGGTCCCTCCCTTTGATGGAGCTGCTGTTGATGCTATTGAC 358  
DB 677 CGGCAACCCCTTTGGTCCCTCCCTTTGATGGAGCTGCTGTTGATGCTATTGAC 736  
QY 359 TCTATTAAATCTTGCACTGCA 380  
DB 737 TCTATTAAATCTTGCACTGCA 758  
RESULT 9  
AG102951      683 bp      DNA      linear      GSS 03-NOV-2001  
LOCUS      Pan troglodytes DNA, clone: PNB-106G16.F, genomic survey sequence.  
DEFINITION      AG102951.1 GI:16723468  
ACCESSION      AG102951.1 GI:16723468  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM      Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
REFERENCE  
AUTHORS      Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Tokoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE      BAC end sequences of library PTB  
JOURNAL  
Unpublished  
2 (bases 1 to 683)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Tokoki, Y., Watanabe, H. and Sakaki, Y.  
Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimbe@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the Red process and may have higher chance of  
clone tracking errors.  
PRIMERS  
Sequencing: -21M13
```

```
LIBRARY  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
1. 683  
/organism="Pan troglodytes"  
/db_xref="taxon:9598"  
/clone="PTB-106G16.F"  
/sex="male"  
/cell_type="lymphoblast"  
/clone_lib="PTB Chimpanzee Male BAC library"  
BASE COUNT      162 a      186 c      167 g      167 t      1 others  
ORIGIN  
Query Match      68.6%; Score 343.8; DB 17; Length 683;  
Best Local Similarity 89.8%; Pred. No. 7.1e-101;  
Matches 403; Conservative 0; Mismatches 42; Indels 4; Gaps 3;  
QY 52 ACTGAGACAGAGCTAGCTGATTTCTTAGGCTGACTAAGATCCTAAGCTTACTG 111  
DB 108 ATTGAGAGAAAGAGCTAGCTGATTTCTTAGGCTGACTAAGATCCTAAGCTTACTG 167  
QY 112 GAAGTGAACCATCCACCTTTAAACAGGGGCTTGCACTTACTGACACCTGACCAAT 171  
DB 168 GAAGTGAACCATCCACCTTTAAACAGGGGCTTGCACTTACTGACACCTGACCAAT 227  
QY 172 CAGAGAGCTCAATTAATTTAGCAAGAGAGCAAGAGGTTAAAGAAATAGCCATCAT 231  
DB 228 CAGAGAGCTCAATTAATTTAGCAAGAGAGCAAGAGGTTAAAGAAATAGCCATCAT 286  
QY 232 CTATTGCTGAGACAGCAGAGAGGAGCAATGATCGGATTAACCCAGCTTCGAG 291  
DB 287 CTATTGCTGAGACAGCAGAGAGGAGCAATGATCGGATTAACCCAGCTTCGAG 346  
QY 292 CCGGCAAGGAAACCCCTTTGGTCCCTCCCTTTGATGGAGAGCTGTTTCATGCT 351  
DB 347 CCGGCAAGGAAACCCCTTTGGTCCCTCCCTTTGATGGAGAGCTGTTTCATGCT 404  
QY 352 ATTTCACCTAATTAATTTGCACTGCTCTGCTGCTCATGTTTCTTAAGCTTGA 411  
DB 405 ATTTCACCTAATTAATTTGCACTGCTCTGCTGCTCATGTTTCTTAAGCTTGA 464  
QY 412 CTGAGCTTTGCTCGCATCCACCATCTGCTG-TTTCGCGCCACCGAGACCCGCTGA 470  
DB 465 CTGAGCTTTGCTCGCATCCACCATCTGCTGTTTGTTCATTTGACAGACCCGCTGA 524  
QY 471 CTCCTACCTCTCTGATCATCAGAGGTGT 499  
DB 525 CTCCTACCTCTCTGATCATCAGAGGTGT 553  
RESULT 10  
NS5091/c      385 bp      mRNA      linear      EST 28-JAN-1997  
LOCUS      yv43e03.61 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
DEFINITION      IMAGE:245500 3', mRNA sequence.  
ACCESSION      NS5091  
VERSION      NS5091.1 GI:1197970  
KEYWORDS  
SOURCE  
ORGANISM      human  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS      Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins  
, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore  
, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,  
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, E.,  
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)
```


QY	304	ACCCCTTTGGGATCCCTCCCTTTGATGGAGGCTGTGTTTCATGCTATTCACCTAT	363
Db	59	A-CCCTTTGGGATCCCTCCCTTTGATGGAGGCTGTGTTTCATGCTATTCACCTAT	1
RESULT 12			
LOCUS	AW971553	522 bp	mRNA linear EST 01-JUN-2000
DEFINITION	EST183842	MAGE resequences, MAGL Homo sapiens cDNA, mRNA sequence.	
ACCESSION	AW971553		
VERSION	AW971553.1	GI:8161399	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 522)		
	Hedge, P., Qi, R., Abernathy, K., Dhurap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quakenbush, J.		
TITLE	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray		
JOURNAL	Unpublished (2000)		
COMMENT	Contract: John Quakenbush The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 3528 Fax: 301 838 0208 Email: john@eigr.org Plate: 292		
FEATURES	Seq primer: Forward.		
source	Location/Qualifiers		
	1..522		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="MAGE resequences, MAGL"		
	/note="Vector: pBluescript-SKm"		
BASE COUNT	125 a 115 c 134 g 148 t		
ORIGIN			
Query Match	66.3%; Score 332.4; DB 10; Length 522;		
Best Local Similarity	93.0%; Pred. No. 3.1e-97;		
Matches	348; Conservative 0; Mismatches 26; Indels 0; Gaps 0.		
QY	4	TCGGCAACCTCCCAACAGCACTTATGATTTCTCTGTGATGGGGGACTGAGACAG	63
Db	374	TCAGCAACTCAACAGCACTTGGTTCTCTGTGATGGGGGACTGAGACAG	315
QY	64	GACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTTAGCTTGGGAAGTGACAC	123
Db	314	GACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTTAGCTTGGGAAGTGACAC	255
QY	124	ATCCACCTTTAAACAGGGGCTTGAACTTAGCTACACTGACCAATAGAGAGCTAC	183
Db	254	ATCCACCTTTAAACAGGGGCTTGCACTTAGCTACACTGACCAATAGAGAGCTAC	195
QY	184	TAAATGCTAATTAGCAAGACAGAGGTTAAAGAAATAGCCATCATTTATGCTGAG	243
Db	194	TAAATGCTAATTAGCAAGCAAGAGGTTAAAGAAATAGCCATCATTTATGCTGAG	135
QY	244	AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAGTCTTGAGCCGAGCAGGCA	303
Db	134	AGCAGCAGTGGGAGGAGCAAGGATTTGCAATTAACCCAGTCTTGAGCAGCAGGCA	75
QY	304	ACCCCTTTGGGATCCCTCCCTTTGATGGAGGCTGTGTTTCATGCTATTCACCTAT	363
Db	74	ACCGCTTTGGGATCCCTCCCTTTGATGGAGGCTGTGTTTCATGCTATTCACCTAT	15
QY	364	TAAATCTGCACT 377	
Db	14	TAAATCTGCACT 1	

```

RESULT 13
T69704/c      T69704          363 bp   mRNA    linear   EST 07-MAR-1995
DEFINITION    yd13a03.sr1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone
IMAGE:66988 3', mRNA sequence.
ACCESSION     T69704
VERSION       T69704.1 GI:680852
KEYWORDS      EST..
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE     1 (bases 1 to 363)
AUTHORS       Hillier L., Clark N., Dubque,T., Elliston,K., Hawkins,M., Holman
              M., Hultman,M., Kucaba,T., Le,M., Lemon,G., Marra,M., Parsons,J.,
              Ritkin,L., Roifling,T., Soares,M., Tan,F., Trevaekis,E., Waterston
              R., Williamson,A., Wohlmann,P. and Wilson,R.
              The Mashu-Merck EST Project
              Unpublished (1995)
TITLE         Washington University School of Medicine
JOURNAL       4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
COMMENT       Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              Insert Size: 766
              High gality sequence strops: 341 Source: IMAGE Consortium, LNL. This
              clone is available royalty-free through LNL / contact the IMAGE
              Consortium (info@image.lnl.gov) for further information.
              Insert Length: 766 Std Error: 0.00
              Seq primer: -2lmj3
              High quality sequence stop: 341.

FEATURES             location/qualifiers
     source           1..363
                     /organism="Homo sapiens"
                     /db_xref="GDB:463693"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:66988"
                     /clone_lib="Soares fetal liver spleen lNFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pTZ19D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I ~ oligo(drr) primer
[5', AACTGGAAAGATTAATTAAAGATCTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptor
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pTZ19 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT      83 a      83 c      87 g      107 t      3 others
ORIGIN
Query Match          65.7%; Score 329; DB 14; Length 363;
Best Local Similarity 97.3%; Pred. No. 3.6e-96;
Matches 354; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

QY      17 CCAACAGCAGCTTAGGTTCCTTCCTTGGAATGGGGAGCTGAGAGAAGAGACTAAGCTGCATTT 76
Db      363 CCAAAGCGACTTAGGTTCCTTCCTTGGAATGGGGAGCTGAGAGAAGAGACTAAGCTGCATTT 304
QY      77 TCCTTAGGCTGACTTAAGAAATCCCTAAGCCTAGCTGAGGAAAGGTACCACACATCCACTTTAAA 136
Db      303 TCCTTAGGCTGACTTAAGAAATCCCTAAGCCTAGCTGAGGAAAGGTACCACACATCCACTTTAAA 244
QY      137 CAAGGAGCTTGCAGACTTAGCTCACACTGACCATCAGAGAGCTCACTAAAAATGTCTAAATT 196
Db      243 CAAGGAGCTTGCAGACTTAGCTCACACTGACCATCAGAGAGCTCACTAAAAATGTCTAAATT 184
QY      197 AGGCCAAGACAGGAGGTAAAGAAATAGCCAATCATCTATTGCTTGAGAGCACAGCAGGAG 256

```

Db 183 AGGCAAGACAGAGGTAAAGAAATACCATCTATTCCTTAGAGCAGACAGAG 124
Qy 257 GGACATGATCGGGATATTAACCCAAAGCTTTCAGAGCCGCAAGCAACCCCTTTGGT 316
Db 123 GGACATGATCGGGATATTAACCCAAAGCTTTCGA-NCGCAACGGCAACCCCTTTGGG- 66
Qy 317 CCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTTCACTCTATTAATCTTGAAC 376
Db 65 CCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTTCACTCTATTAATCTTGAAG 6
Qy 377 TGCA 380
Db 5 NACA 2

RESULT 14
LOCUS AA860368/c 342 bp mRNA linear EST 31-DEC-1998
DEFINITION aj59c05.a1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1394600
3', similar to contains PTR7.t1 PTR5 repetitive element ;, mRNA
sequence.
ACCESSION AA860368
VERSION AA860368.1 GI:2954363
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 342)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL ac:
www-bio.lnl.gov/bbtp/image/image.html
Insert Length: 1305 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 319.
Location/Qualifiers
1. .342
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394600"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5 and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 75 a 78 c 82 g 107 t
ORIGIN

Query Match 65.2%; Score 326.6; DB 9; Length 342;
Best Local Similarity 98.8%; Pred. No. 2.2e-95;
Matches 329; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 48 GGGGACTGAGAGACGAGCTAGCTGATTTCTTAGGCTGACTTAAGATCCCTTAAGCTTAG 107
Db 342 GGGGACTGAGAGACGAGCTAGCTGATTTCTTAGGCTGACTTAAGATCCCTTAAGCTTAG 283

Qy 108 CTGGAGGTGACCAATCCACTTTTAAACAAGGGCTTTCAGACTTAAGCTCAACTGAC 167
Db 282 CTGGGAAGTGAACCAATCCACTTTTAAACAAGGGCTTTCAGACTTAAGCTCAACTGAC 223
Qy 168 CAATCAGAGACTCACTAAATGCTTAATTAGCAAAAGCAGAGGTAAAGAAATAGCCAA 227
Db 222 CAATCAGAGACTCACTAAATGCTTAATTAGCAAAAGCAGAGGTAAAGAAATAGCCAA 163
Qy 228 TCATCTATTGCTGAGACAGAGGAGGACATGATCGGGATATTAACCCAAAGCTT 287
Db 162 TCATCTATTGCTGAGACAGAGGAGGACATGATCGGGATATTAACCCAAAGCTT 103
Qy 288 CGAGCCGCAACGGCAACCCCTTTGGGCTCCCTTCCTTTGATGAGAGCTCTGTTTCA 347
Db 102 CGAGCCGCAACGGCAACCCCTTTGGGCTCCCTTCCTTTGATGAGAGCTCTGTTTCA 43
Qy 348 TGCTATTTCACCTTATTAATCTTGAACCTGCA 380
Db 42 TGCTATTTCACCTTATTAATCTTGAACCTGCA 10

RESULT 15
LOCUS R27412/c 424 bp mRNA linear EST 24-APR-1995
DEFINITION yH46d11.a1 Soares Placenta Mb2HP Homo sapiens cDNA clone
IMAGE:132789 3', mRNA sequence.
ACCESSION R27412
VERSION R27412.1 GI:783547
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 424)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterston
,R., Williamson,A., Woldmann,P. and Wilson,R.
TITLE The Washu-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 695
High quality sequence stops: 367
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 695 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 367.
Location/Qualifiers
1. .424
/organism="Homo sapiens"
/db_xref="GDB:538431"
/db_xref="taxon:9606"
/clone="IMAGE:132789"
/clone_lib="Soares placenta Mb2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
ACCTGGAAGATTCGGCGCGGAGGATTTTCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library

BASE COUNT 92 a constructed by Bento Soares and M.Fatima Bonaldo. "
ORIGIN 96 c 115 g 121 t

Query Match 65.1%; Score 326.1; DB 14; Length 424;
Best Local Similarity 93.9%; Pred. No. 118e-95;
Matches 351; Conservative 0; Mismatches 21; Indels 2; Gaps 1;

QY 4 TCGGCCAACCTGCCCAAGACAGCACTTAGGTTTCCGTGAGATGGGGAGCTGAGACAG 63
DB 373 TCGATCAACCTGCCCAAGACAGCACTTAGGTTTCCGTGAGATGGGGAGCTGAGACAG 314
QY 64 GACTAGCTGATTTCTAGGCTGACTAAGATCCCTAAGCTAGCTGGAGGTGACCAAC 123
DB 313 GACTAGCTGATTTCTAGGCTGACTAAGATCCCTAAGCTAGCTGGAGGTGACCTGC 254
QY 124 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTGACCACTGACCAATCAGAGAGCTGAC 183
DB 253 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTGACCACTGACCAATCAGAGAGCTGAC 196
QY 184 TAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG 243
DB 195 TAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG 136
QY 244 AGCAGAGAGAGGAGCAATGATGGGATATTAACCTCAAGCTTGAAGCCGCAACGGCA 303
DB 135 AGCAGAGAGAGGAGGCAAGATCGGATATTAACCTCAAGCTTGAAGCCGCAACGGCA 76
QY 304 ACCCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTTGTTTTCATGCTATTGCACTTAT 363
DB 75 AACCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTTGTTTTCATGCTATTGCACTTAT 16
QY 364 TAAATCTTGCAACT 377
DB 15 TAAATCTTGCAACT 2

Search completed: April 19, 2003, 14:46:37
Job time : 1271.51 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 17, 2003, 04:28:30 ; Search time 1703.47 Seconds

(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 1 gtcgcggcgaaccccccacccca.....ctggatcatgcagggtgtcc 501

Sequence: IDENTITY NUC
Gapop 10.0 , Gapept 1.0

Scoring table: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

Genbank1.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_com.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_scs.*
28: em_un.*
29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rtd.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hcg_hum.*
40: em_hcg_mus.*
41: em_hcg_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	10499	6 AX007980	AX007980 Sequence
2	498	99.4	56093	6 AX329572	AX329572 Sequence
3	498	99.4	56093	9 HSAC000064	AC000064 Human BAC
4	498	99.4	149194	9 AC007566	AC007566 Homo sapi
5	450	89.8	711	6 AX007998	AX007998 Sequence
6	434	86.6	174019	9 AP001538	AP001538 Homo sapi
7	434	86.6	340000	9 AP001674	AP001674 Homo sapi
8	431.2	86.1	163803	9 AC093531	AC093531 Homo sapi
9	426	85.0	839	9 AF127226	AF127226 Homo sapi
10	426	85.0	1329	6 AX001030	AX001030 Sequence
11	425.6	85.0	783	6 AX000970	AX000970 Sequence
12	415.6	83.0	251124	2 AC092510	AC092510 Papio cyn
13	409.6	81.8	187321	2 AC092843	AC092843 Homo sapi
14	409	81.6	83412	9 AC092843	AC092843 Homo sapi
15	407.2	81.3	140756	2 AL133038	AL133038 Human DNA
16	407.2	81.3	183499	2 AL607153	AL607153 Homo sapi
17	400.6	80.0	169029	2 CNS06C7R	AL339080 Human chr
18	400.2	79.9	110000	2 AL353584	AL353584 Homo sapi
19	400.2	79.9	170746	2 AL392173	AL392173 Human DNA
20	398.8	79.6	191863	9 AC010888	AC010888 Homo sapi
21	396.6	79.2	88328	9 AL357874	AL357874 Human DNA
22	396.4	79.1	157711	9 AC055738	AC055738 Homo sapi
23	396.4	79.1	207945	2 AC021652	AC021652 Homo sapi
24	396.4	79.1	318442	2 AC021652	AC021652 Homo sapi
25	396.2	79.1	711	6 AX007997	AX007997 Sequence
26	392.8	78.4	114455	9 AC093742	AC093742 Homo sapi
27	391.4	78.1	165059	2 AC106856	AC106856 Homo sapi
28	390	77.8	137492	2 AL161721	AL161721 Homo sapi
29	389.8	77.8	181753	9 AC124075	AC124075 Homo sapi
30	389.6	77.8	143590	2 HS836J3	AL035706 Human DNA
31	388	77.4	119406	2 AC103595	AC103595 Homo sapi
32	387.4	77.3	76169	9 AC003014	AC003014 Human PAC
33	386.8	77.2	180635	9 AL360169	AL360169 Human DNA
34	386	77.0	145785	2 AC016173	AC016173 Homo sapi
35	385.8	77.0	161049	2 AL591842	AL591842 Homo sapi
36	385.4	76.9	111079	9 HS419C19	AL035407 Human DNA
37	385.2	76.9	156177	9 AC093896	AC093896 Homo sapi
38	384.8	76.8	42216	9 AC092899	AC092899 Homo sapi
39	384.8	76.8	170414	2 AC026019	AC026019 Homo sapi
40	384.6	76.8	119481	9 AL136234	AL136234 Human DNA
41	384.4	76.7	91097	9 AP001600	AP001600 Homo sapi
42	384.4	76.7	118241	9 AP001599	AP001599 Homo sapi
43	384.4	76.7	185018	9 AC110614	AC110614 Homo sapi
44	384.4	76.7	340000	9 AP001697	AP001697 Homo sapi
45	383	76.4	181850	9 AC018680	AC018680 Homo sapi

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE
AX007980	Sequence 3 from Patent WO9967395.	AX007980	AX007980.1	GI:9995677	human.	Homo sapiens	1 (bases 1 to 10499)	Perin,J.P., Rieger,F. and Alliel,P.M.	endogenous retroviral motifs, and their uses
AX007980	Sequence 3 from Patent WO9967395.	AX007980	AX007980.1	GI:9995677	human.	Homo sapiens	1 (bases 1 to 10499)	Perin,J.P., Rieger,F. and Alliel,P.M.	endogenous retroviral motifs, and their uses

JOURNAL Patent: WO 967395-A 3 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES Location/Qualifiers
SOURCE 1.10499
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 2455 t
ORIGIN

Query Match 100.0%; Score 501; DB 6; Length 10499;
Best Local Similarity 100.0%; Pred. No. 6.2e-151;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCTCGGCAACCTCCCAACAGCACTTGTGTTTCTGTGAGATGGGGAGCTAGAGA 60
DB 9500 GCTCGGCAACCTCCCAACAGCACTTGTGTTTCTGTGAGATGGGGAGCTAGAGA 9559
OY 61 CAGGACTAGCTGATTTCTTGAAGATTCCTAAGCTTGAAGAGGTGAC 120
DB 9560 CAGGACTAGCTGATTTCTTGAAGATTCCTAAGCTTGAAGAGGTGAC 9619
OY 121 CAGATCCACCTTTAAACAGGGGCTTGCACTTGTCACTGACCAATGAGAGCT 180
DB 9620 CAGATCCACCTTTAAACAGGGGCTTGCACTTGTCACTGACCAATGAGAGCT 9679
OY 181 CACTAAATGCTAATTAGGCAAGAGAGGTAAAGAAATAGCAATCATCTATTGCT 240
DB 9680 CACTAAATGCTAATTAGGCAAGAGAGGTAAAGAAATAGCAATCATCTATTGCT 9739
OY 241 GAGACACAGCAGAGAGGCAATGATCGGATATTAACCAAGCTTCGAGCCGCAACG 300
DB 9740 GAGACACAGCAGAGAGGCAATGATCGGATATTAACCAAGCTTCGAGCCGCAACG 9799
OY 301 GGAACCCCTTTGGGTCCTCCCTTTGATAGGAGTCTGTGTTTCAGCTATTCACTC 360
DB 9800 GGAACCCCTTTGGGTCCTCCCTTTGATAGGAGTCTGTGTTTCAGCTATTCACTC 9859
OY 361 TATTAAATCTTCACTGACTCTTGTGTCATGTTCTTACGGCTTGAAGCTTGA 420
DB 9860 TATTAAATCTTCACTGACTCTTGTGTCATGTTCTTACGGCTTGAAGCTTGA 9919
OY 421 CGCTGGCCATCCACCACTGCTGTTGCGCCACCGGAGACCCGCGCTGATCCCATCC 480
DB 9920 CGCTGGCCATCCACCACTGCTGTTGCGCCACCGGAGACCCGCGCTGATCCCATCC 9979
OY 481 TCTGATCATGCAAGGTGTC 501
DB 9980 TCTGATCATGCAAGGTGTC 10000

RESULT 2
AX329572 56093 bp DNA linear PAT 09-JAN-2002
LOCUS AX329572
DEFINITION Sequence 81 from Patent WO0194629.
ACCESSION AX329572
VERSION AX329572.1 GI:18102550
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ehner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;
Avalon Pharmaceuticals (US)

FEATURES
SOURCE location/Qualifiers
1. 56093
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match 99.4%; Score 498; DB 6; Length 56093;
Best Local Similarity 100.0%; Pred. No. 7.2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TCGGCAACCTCCCAACAGCACTTGTGTTTCTGTGAGATGGGGAGCTAGAGA 63
DB 37504 TCGGCAACCTCCCAACAGCACTTGTGTTTCTGTGAGATGGGGAGCTAGAGA 37563
OY 64 GACTAGCTGATTTCTTGAAGATTCCTAAGCTTGAAGAGGTGAC 123
DB 37564 GACTAGCTGATTTCTTGAAGATTCCTAAGCTTGAAGAGGTGAC 37623
OY 124 ATCCACTTTAAACAGGGGCTTGCACTTGTCACTGACCAATCAGAGAGCTAC 183
DB 37624 ATCCACTTTAAACAGGGGCTTGCACTTGTCACTGACCAATCAGAGAGCTAC 37683
OY 184 TAAATGCTAATTAGGCAAGAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG 243
DB 37684 TAAATGCTAATTAGGCAAGAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG 37743
OY 244 AGCAGCAGAGAGGCAATATCGGATATTAACCAAGCTTCGAGCCGCAACGCA 303
DB 37744 AGCAGCAGAGAGGCAATATCGGATATTAACCAAGCTTCGAGCCGCAACGCA 37803
OY 304 ACCCCCTTTGGGTCCTCCCTTTGATAGGAGCTGTGTTTCAGTATTCACTCTAT 363
DB 37804 ACCCCCTTTGGGTCCTCCCTTTGATAGGAGCTGTGTTTCAGTATTCACTCTAT 37863
OY 364 TAAATCTTGAACCTGACTCTTGTGTCATGTTCTTACGGCTTGAAGCTTGA 423
DB 37864 TAAATCTTGAACCTGACTCTTGTGTCATGTTCTTACGGCTTGAAGCTTGA 37923
OY 424 TCGCATCACCACTGCTGTTTGGCGCAACCGGCGCTGATCCCATCCCTCT 483
DB 37924 TCGCATCACCACTGCTGTTTGGCGCAACCGGCGCTGATCCCATCCCTCT 37983
OY 484 GGATCATGCAAGGTGTC 501
DB 37984 GGATCATGCAAGGTGTC 38001

RESULT 3
HSAC000064 56093 bp DNA linear PRI 13-NOV-1996
LOCUS HSAC000064
DEFINITION Human BAC clone RG083M05 from 7q21-7q22, complete sequence.
ACCESSION AC000064
VERSION AC000064.1 GI:1669369
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 56093)
AUTHORS Pauley, A.
TITLE The sequence of H. sapiens BAC clone RG083M05
JOURNAL unpublished (1996)
REFERENCE 2 (bases 1 to 56093)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1996)
COMMENT Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
e-mail: saplens@wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
sections once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The
library contains cloned DNA from a human male fibroblast cell line
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBET0

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this
clone is at base position 1 of H_RG083M05; actual end is at 56093
of H_RG083M05

This clone contains STS SWS1725.

FEATURES

SOURCE

location/Qualifiers

1..56093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"

/clone_id="CTTB-978SK-B"
complement(838..1131)
/rpt_family="ALU"

repeat_region

/gene="WUGSC:H_RG083M05.1"

CDS

join(<1360..1503,4181..4370,4587..4774,6422..6556,
9483..9547,11631..11773,11864..12021,13131..13296,
14885..14988,16349..16546,16837..16971)
/gene="WUGSC:H_RG083M05.1"
/note="ATPase; strong similarity to peroxisome
biosynthesis protein Pasi (PID:g1172019); coded for by
human CDNA C04279 (NID:g146530)"

/codon_start=1
/protein_id="AAB46346.1"
/db_xref="GI:1669371"

/translation="KRLNIQKLEVAFAVMNQSVLLDDLLIAGLPAVEHEH
SPDAORCEILCNVKNKLDGDKFTDLOHAKETGKVADFTLVRAHLSRL
SRQISTRREKVLTTDPOKALRGTPSLSSVNHKRGDGMKIGLHVRLIWD
TTQPAKVCLEKEKPELFLANLPTRQRTGILLYGPDTGKTLGAVIARBRNFTISV
KGPBLISKYIGASQAVADIFIRQAAPCILFDEFESIAPRGHDTVTVDVNG
LITQDVGVEGLQGVYVLAATSRPLIDPALRPGLDCCVCPDPDQVTTLSLEKTKQ
QMLHSFLVSRLEINLVLSDSLPLADVDVLOHVASTDSGADIKALINQALALHG
MLLSKMSRIIPDESKENMYRLYFGSSVSEELGNGTSSDLSGCLSAPESTMQDLPV
GKQDLFSQPYVLRTRASQGCCELVQEQKQDQRADISITIKGYRSGGDESMQGPRI
KTRLAISQSHLMTALGHTRPISISDDMKNFVEL"

repeat_region

/rpt_family="ALU"

repeat_region

complement(6581..7133)

repeat_region

/rpt_family="ALU"

repeat_region

complement(7767..8037)

repeat_region

/rpt_family="ALU"

repeat_region

complement(8186..8472)

repeat_region

/rpt_family="ALU"

repeat_region

8473..8625

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 3' EST H75782 (NID:g1049794), bases
287-444"

repeat_region

8841..9161

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 5' EST H75921 (NID:g1050050), bases
21-348"

repeat_region
complement(12612..12907)

/rpt_family="ALU"

misc_feature

/gene="WUGSC:H_RG083M05.1"

repeat_region

13794..13877

misc_feature

/gene="WUGSC:H_RG083M05.1"

repeat_region

13878..13906

repeat_region

/rpt_family="ALU"

repeat_region

complement(14110..14137)

repeat_region

/rpt_family="ALU"

repeat_region

17227..17522

misc_feature

/rpt_family="ALU"

repeat_region

18667..19235

repeat_region

/note="match to human fetal brain 5' EST D61494
(NID:g970409), bases 1-255, and to human 3' EST R07476
(NID:g759399)"

repeat_region

19550..19670

misc_feature

/rpt_family="ALU"

repeat_region

21507..37303

exon

/note="similarity to various SS-RNA virus polypeptides;
pseudogene; region of matches and close matches to
multiple human ESTs, see R68740 (NID:g842257)"

repeat_region

37316..37489

repeat_region

/note="Grail prediction, score = 80"

repeat_region

/evidence=not experimental

repeat_region

complement(38938..39224)

repeat_region

/rpt_family="ALU"

repeat_region

39225..39707

repeat_region

/note="match to multiple human ESTs, see N30113
(NID:g1148633)"

repeat_region

39800..40085

repeat_region

/rpt_family="ALU"

repeat_region

complement(40247..40538)

repeat_region

/rpt_family="ALU"

repeat_region

complement(40632..40924)

repeat_region

/rpt_family="ALU"

repeat_region

complement(42283..42891)

repeat_region

/rpt_family="ALU"

repeat_region

complement(44574..45613)

repeat_region

/rpt_family="ALU"

repeat_region

complement(45614..45737)

repeat_region

/note="match to human 3' EST H48898 (NID:g988738), bases
129-333"

repeat_region

complement(46107..47026)

repeat_region

/note="match to multiple human ESTs, see N81064
(NID:g1243765), H48897 (NID:g988737), and W78831
(NID:g273146)"

repeat_region

complement(47027..47318)

repeat_region

/rpt_family="ALU"

repeat_region

complement(47365..47782)

repeat_region

/note="match to multiple human ESTs, see W37495
(NID:g1319089)"

repeat_region

47898..48115

repeat_region

/note="match to human 5' EST H62306 (NID:g1015138), bases
93-368"

repeat_region

complement(48116..48405)

repeat_region

/rpt_family="ALU"

repeat_region

complement(48406..48584)

repeat_region

/note="match to human 3' EST N29952 (NID:g1148472), bases
290-455, and 5' EST R12730 (NID:g765806)"

repeat_region

complement(48787..49405)

repeat_region

/rpt_family="ALU"

repeat_region

complement(49406..49534)

repeat_region

/note="match to human 3' EST R65794 (NID:g838432), bases
309-440"

```

repeat_region      complement(49638..49672)
                    /rpt_family="ALU"
misc_feature        complement(49674..49896)
                    /note="match to human 3' EST N29952 (NID:g1148472) and 5'
                    EST N29938 (NID:g1148458), sequences are from opposite
                    ends of the same clone"
gene                complement(49698..51806)
                    /gene="WUGSC:H.RG083M05.2"
                    complement(join(49698..49888,51575..51806))
                    /gene="WUGSC:H.RG083M05.2"
                    /note="coded for by human cDNA M37389 (NID:g119205),
                    R65891 (NID:g838529), R65794 (NID:g838432) and R65794
                    (NID:g838432)"
                    /protein_id="AAB46345.1"
                    /codon_start=1
                    /db_xref="GI:1669370"
                    /translation="MFYFPOCGT::FCPGVYVQIGNVVIDEODKPYAIGRP
                    IODYCEKSAALTWLITPLSPRODIPDASTITIGBEDLPKMEYLEFVCHAFSEYFK
                    SRSSPFPVTPRPEKGYIWTIVGP:PALTIKESVANHL"
exon                complement(51576..51758)
                    /gene="WUGSC:H.RG083M05.2"
                    /note="trail prediction, score = 86"
repeat_region      /evidence=not_experimental
                    complement(52052..52329)
                    /rpt_family="h1"
                    /note="match to human EST M79192 (NID:g273505) base 2-289"
misc_feature        55557..55843
Query Match        99.4%; Score 498; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. No. 7.2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
4  TCGGCAACCTCCCAACAGCACTAGGTTTCCTGTTGAGATGGGGAGCAGAGACAG 63
Db  37504 TCGGCAACCTCCCAACAGCACTAGGTTTCCTGTTGAGATGGGGAGCAGAGACAG 37563
Qy  64  GACTAGTGGATTTCCTAGGCTGACTAAGAAATCCCTAGCTGGAGGTGACAC 123
Db  37564 GACTAGTGGATTTCCTAGGCTGACTAAGAAATCCCTAGCTGGAGGTGACAC 37623
Qy  124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCA(ACCTGACCAATAGAGAGCTCAC 183
Db  37624 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCA(ACCTGACCAATAGAGAGCTCAC 37683
Qy  184 TAAATGCTAATTAGCAAGAGAGAGGTAAAGAAATAGCCATCATCTATTGCTGAG 243
Db  37684 TAAATGCTAATTAGCAAGAGAGAGGTAAAGAAATAGCCATCATCTATTGCTGAG 37743
Qy  244 AGCAGCAGAGAGAGGACATGATCGGAGATATAACCCCAAGTCTTGAGCCGGCAACGGCA 303
Db  37744 AGCAGCAGAGAGAGGACATGATCGGAGATATAACCCCAAGTCTTGAGCCGGCAACGGCA 37803
Qy  304 ACCCCCTTTGGGTCCTCCCTTTGATATGGAGCTCTATTTCATGCTATTTCACCTCAT 363
Db  37804 ACCCCCTTTGGGTCCTCCCTTTGATATGGAGCTCTATTTCATGCTATTTCACCTCAT 37863
Qy  364 TAAATCTTGCACTGACTCTTCTGCTCATGCTTTCTACGGCTTGAGGCTGAGCTTTGCG 423
Db  37864 TAAATCTTGCACTGACTCTTCTGCTCATGCTTTCTACGGCTTGAGGCTGAGCTTTGCG 37923
Qy  424 TCGCATCCACCACTGCTGTTTGCAGCAACGAGACGCGCGCTGACTCCCATCCCTCT 483
Db  37924 TCGCATCCACCACTGCTGTTTGCAGCAACGAGACGCGCGCTGACTCCCATCCCTCT 37983
Qy  484 GGATCATGCAAGGCTGCC 501
Db  37984 GGATCATGCAAGGCTGCC 38001

```

```

RESULT 4
AC007566/c AC007566 149194 bp DNA linear PRI 01-MAR-2002
LOCUS      Homo sapiens BAC clone CTB-1005 from 7q21-7q22, complete sequence.
DEFINITION

```

```

ACCESSION AC007566
VERSION   AC007566.2
KEYWORDS  GI:11181861
SOURCE    HTG.
ORGANISM  Homo sapiens.
           Homo sapiens.
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS   1 (bases 1 to 149194)
           Sulston, J.E. and Waterston, R.
TITLE     Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
PUBMED    9847074
REFERENCE  2 (bases 1 to 149194)
AUTHORS   Du, Z.
TITLE     The sequence of Homo sapiens BAC clone CTB-1005
JOURNAL   Unpublished (2001)
REFERENCE  3 (bases 1 to 149194)
AUTHORS   Waterston, R.H.
TITLE     Direct Submission
JOURNAL   Submitted (15-MAY-1999) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  4 (bases 1 to 149194)
AUTHORS   Waterston, R.
TITLE     Direct Submission
JOURNAL   Submitted (02-OCT-2000) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE  5 (bases 1 to 149194)
AUTHORS   Waterston, R.H.
TITLE     Direct Submission
JOURNAL   Submitted (16-NOV-2000) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  6 (bases 1 to 149194)
AUTHORS   Waterston, R.H.
TITLE     Direct Submission
JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  7 (bases 1 to 149194)
AUTHORS   Waterston, R.H.
TITLE     Direct Submission
JOURNAL   Submitted (06-FEB-2002) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
REFERENCE  8 (bases 1 to 149194)
AUTHORS   Waterston, R.
TITLE     Direct Submission
JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On Nov 16, 2000 this sequence version replaced gi:4835815.
COMMENT   ----- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc
           Contact: sapiens@watson.wustl.edu
           ----- Summary Statistics
           Center project name: H_RG010605

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7>, send <mailto:malco@egreen.nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.reagen.com>).

VECTOR: pBel0BAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-91H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

SOURCE

Location/Qualifiers
1. 149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_lib="CTB-978SK-B"
1. 1634
/rpt_family="L1"
repeat_region
2248..2388
/note="match to EST BG752883 (NID:g14063536)"
misc_feature
2248..2387
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
2248..2387
/note="match to EST AW579261 (NID:g7254310)"
misc_feature
2248..2387
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
2248..2387
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
2248..2374
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
2248..2287
/note="match to EST AV686676 (NID:g10288539)"
2253..2387
/note="match to EST AU123510 (NID:g10948226)"
2344..2387
/note="match to EST AV686676 (NID:g10288539)"
2696..3066
/rpt_family="MaLR"
3108..3392
/rpt_family="Alu"
3540..3628
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
3540..3628
/note="similar to Mus musculus EST BE253526 (NID:g8946272)"
3542..3632
/note="match to EST AV686676 (NID:g10288539)"
3542..3628
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
3542..3628
/note="match to EST AU123510 (NID:g10948226)"
3542..3628
/note="match to EST BG766882 (NID:g14077535)"

3542..3628
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
3542..3609
/note="match to EST AW579261 (NID:g7254310)"
misc_feature
3562..3628
/note="match to EST BE272564 (NID:g9146913)"
misc_feature
3598..3628
/note="match to EST BE299708 (NID:g9183456)"
misc_feature
3714..3785
/note="match to EST BG260659 (NID:g12770475)"
misc_feature
3717..3785
/note="similar to Mus musculus EST BE94936 (NID:g10678674)"
3719..3785
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
3719..3785
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature
3719..3785
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
3719..3785
/note="match to EST BE272564 (NID:g9146913)"
misc_feature
3719..3785
/note="match to EST BE299708 (NID:g9183456)"
misc_feature
3719..3785
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
3719..3785
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
3719..3785
/note="similar to Mus musculus EST BE253526 (NID:g8946272)"
3860..3862
/note="match to EST AV686676 (NID:g10288539)"
4730..4827
/rpt_family="L2"
4834..4925
/rpt_family="L1"
5715..5860
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
5715..5860
/note="match to EST AU123510 (NID:g10948226)"
5715..5860
/note="match to EST BE272564 (NID:g9146913)"
5715..5860
/note="match to EST BE299708 (NID:g9183456)"
5715..5860
/note="match to EST BG260659 (NID:g12770475)"
5715..5860
/note="match to EST BG766882 (NID:g14077535)"
5715..5860
/note="match to EST BI160365 (NID:g14620366)"
5715..5860
/note="similar to Mus musculus EST BE94936 (NID:g10678674)"
Query Match 99.4%; Score 498; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 8,2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTTAGCTTTCTGTTGAGATGGGGACCTGAGACAG 63
DB 84500 TCGGCAACCTCCCAACAGCACTTAGCTTTCTGTTGAGATGGGGACCTGAGACAG 84441
QY 64 GACTAGCTGATTTCTAGGCTGACTAAGATCCCTAGGCTAGCTGGAGAGTGACAC 123
DB 84440 GACTAGCTGATTTCTAGGCTGACTAAGATCCCTAGGCTAGCTGGAGAGTGACAC 84381
QY 124 ATCCACCTTTAAACAGGGGCTTGCAACTTAGCTACACCTGACCAATCAGAGAGCTCAC 183
DB 84380 ATCCACCTTTAAACAGGGGCTTGCAACTTAGCTACACCTGACCAATCAGAGAGCTCAC 84321
QY 184 TAAATGCTAATTGGCAAGACGAGGTTAAAGAAATAGCCATCATCTTTGCTCGAG 243
DB 84320 TAAATGCTAATTGGCAAGACGAGGTTAAAGAAATAGCCATCATCTTTGCTCGAG 84261

QY 244 AGCAGCAGAGAGGAGCAATATCGGAGATATTAAC CCAATGCTTGACCGCGCAACGCA 303
Db 84260 AGCAGCAGAGAGGAGCAATATCGGAGATATTAAC CCAATGCTTGACCGCGCAACGCA 84201
QY 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCACTCTAT 363
Db 84200 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCACTCTAT 84141
QY 364 TAAATCTTGCAACTGACTCTTCTGATCCATGTTT TTAACGCTTGAGCTGAGCTTTGCG 423
Db 84140 TAAATCTTGCAACTGACTCTTCTGATCCATGTTT TTAACGCTTGAGCTGAGCTTTGCG 84081
QY 424 TCGCCATCCACCACTGCTGTTTGGCCGCAACGCAACCGCGCTGACTCCATCCCTCT 483
Db 84080 TCGCCATCCACCACTGCTGTTTGGCCGCAACGCAACCGCGCTGACTCCATCCCTCT 84021
QY 484 GGATCATCGAGGTGTCC 501
Db 84020 GGATCATCGAGGTGTCC 84003

RESULT 5

AX007998
LOCUS AX007998 711 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 21 from Patent WO967395.
ACCESSION AX007998
VERSION AX007998.1 GI:995695
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
TITLE Perin, J.P., Rieger, F. and Alliel, P.M.
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses

JOURNAL
TITLE INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)

FEATURES
source location/Qualifiers
1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 171 a 209 c 160 g 171 t
ORIGIN

Query Match 89.8%; Score 450; LB 6; Length 711;

Best Local Similarity 100.0%; Pred. No. 1.6e-14; Indels 0; Gaps 0;
Matches 450; Conservative 0; Mismatches 0;

QY 52 ACTGAGAGCAGAGCTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGG 111
Db 1 ACTGAGAGCAGAGCTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGG 60
QY 112 GAAGTGACACATCCACCTTTAAACGCGGGCTTGCACCTTAGCTCACACCTGACCAAT 171
Db 61 GAAGTGACACATCCACCTTTAAACGCGGGCTTGCACCTTAGCTCACACCTGACCAAT 120
QY 172 CAGAGAGCTCATTAATGCTAATTAAGCAAGAGAGTAAAGAAATTAACCAATCT 231
Db 121 CAGAGAGCTCATTAATGCTAATTAAGCAAGAGAGTAAAGAAATTAACCAATCT 180
QY 232 CTAATGCTGAGAGCAGAGAGAGGAGCAATGATCGAGATATTAACCAAGCTTTGAG 291
Db 181 CTAATGCTGAGAGCAGAGAGAGGAGCAATGATCGAGATATTAACCAAGCTTTGAG 240
QY 292 CCGGCAACGCAACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCACTGCT 351
Db 241 CCGGCAACGCAACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCACTGCT 300
QY 352 ATTTCATCTATTAATCTTTCGACTGCACTCTTCTGCTCATGTTTCTTACGGCTTGA 411
Db 301 ATTTCATCTATTAATCTTTCGACTGCACTCTTCTGCTCATGTTTCTTACGGCTTGA 360

QY 412 CTGAGCTTTCGCTCGCATTCACCACTGCTGTTTCCGCGCAACGCGAGACCCGCTGAC 471
Db 361 CTGAGCTTTCGCTCGCATTCACCACTGCTGTTTCCGCGCAACGCGAGACCCGCTGAC 420
QY 472 TCCCATCCCTTGATCATGAGAGGTGTCC 501
Db 421 TCCCATCCCTTGATCATGAGAGGTGTCC 450

RESULT 6

AP001538/c 174019 bp DNA linear PRI 25-MAR-2000
LOCUS AP001538
DEFINITION Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B853K11,
L156-APP region, complete sequence.
ACCESSION AP001538
VERSION AP001538.1 GI:7328982
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:B853K11.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
TITLE Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

JOURNAL
TITLE Published Only in Database (2000)
2 (bases 1 to 174019)

REFERENCE
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

JOURNAL
TITLE Direct Submission
Submitted (23-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kilaseato Univ., 1-15-1 Kilaseato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp, Tel:81-42-778-9923,
Fax:81-42-778-9924)

FEATURES
source location/Qualifiers
1..174019
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1-q21.2"
/clone="B853K11"

BASE COUNT 53303 a 31416 c 32248 g 57052 t
ORIGIN

Query Match 86.6%; Score 434; DB 9; Length 174019;

Best Local Similarity 92.0%; Pred. No. 4.9e-129; Indels 0; Gaps 0;
Matches 458; Conservative 0; Mismatches 40;

QY 4 TCGGCAACCTCCCAACAGACACTTAGGTTTCTGTTGAGATGGGAGCTGAGAGACAG 63
Db 36468 TTGGCAACCTCCCAACAGACACTTAGGTTTCTGTTGAGAGGCGGAGTACGAGAGACAG 36409
QY 64 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAAGGTGACCA 123
Db 36408 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAAGGTGACCA 36349
QY 124 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAATCAAGAGCTTAC 183
Db 36348 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAATCAAGAGCTTAC 36289
QY 184 TAAATGCTAATTAAGCAAGAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 243
Db 36288 TAAATGCTAATTAAGCAAGAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 36229
QY 244 AGCAGCAGAGAGGAGCAATATCGGAGATATTAAC CCAATGCTTGACCGCGCAACGCA 303
Db 36228 AGCAGCAGAGAGGAGCAATATCGGAGATATTAAC CCAATGCTTGACCGCGCAACGCA 36169
QY 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCACTCTAT 363

Db 36168 ACCCCCTTGGGTCCCTCCCTTTGATGAGAGCTCTGTTTACTCTATTTCACCTAC 36109
Qy 364 TAAATCTGCAACTGACCTCTCTGTCATGTTTCTTACAGCTTGAGCTGAGCTTTGCG 423
Db 36108 TAAATCTGCAACTGACCTCTCTGTCATGTTTCTTACAGCTTGAGCTGAGCTTTGCG 36049
Qy 424 TCGGCATCCACCACTGCTGTTTGGCGCAGACCGCGGCTGACTCCCATCCCTCT 483
Db 36048 TCGCTTCCACCACTGCTGTTTGGCGCAGACCGCTGACTGACTCCCATCCCTCC 35989
Qy 484 GGATTCATCAGAGGTCTCC 501
Db 35988 GGATTCGCGAGAGCTGCTCC 35971

RESULT 7
LOCUS AP001674/c 340000 bp DNA linear PRI 10-MAY-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 18/105.
ACCESSION AP001674 AL163219 BA000005
VERSION AP001674.1 GI:7768666
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Paterson,D., Reichwald,K., Rump,A., Schilhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,K., Nagamine,K., Mitsuayama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstrek,G.,
Hornischler,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hemmi,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaepo,M.L.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
2 (bases 1 to 340000)
20289799
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Paterson,D., Reichwald,K., Rump,A., Schilhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuayama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstrek,G.,
Hornischler,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hemmi,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaepo,M.L.
Direct Submission
Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
On May 30, 2000 this sequence version replaced gi:7717271.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan.
* e-mail: hattori@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany, *

FEATURES
source
1..340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
1..6149
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R749F18, 5' partial"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AP001537"
1..165231
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R853K11, 5' partial"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AP001538"
136796..315056
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R845I12"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AF245226"
156433..277809
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="P646P19"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AL157359"
207594..250588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="Q3D19"
/clone_1lb="AL121NC02-Q Coesmid library"
/note="Accession No. AL049753"
276958..301481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R755I19"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AP001345"
299634..340000
/organism="Homo sapiens"

```
source
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R28M9_3' partial"
/clone_1db="R28M9-11 BAC library"
/note="Accession No. AP00431"
311751..340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="C103E2331_3' partial"
/clone_1db="ICRF403 Cosmid library"
/note="Accession No. AL32933"
15..444
repeat_region
/note="L1MA4A"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
555..3531
repeat_region
/note="L1MA4A"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
3770..3876
repeat_region
/note="L1"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
complement(5645..6178)
/note="L1MA2"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
6181..6534
repeat_region
/note="THE1B"
/rpc_family="LTR/MaLR"
/rpc_type=DISPERSED
8021..8370
repeat_region
/note="THE1B"
/rpc_family="LTR/MaLR"
/rpc_type=DISPERSED
complement(8373..10032)
/note="L1MA2"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
10213..10520
repeat_region
/note="Alu"
/rpc_family="SINE/Alu"
/rpc_type=DISPERSED
15810..15851
repeat_region
/note="(CA)n"
/rpc_family="Simple_repeat"
/rpc_type=TANDEM
complement(16628..16697)
/note="MERS3"
/rpc_family="DNA"
/rpc_type=DISPERSED
16638..16715
repeat_region
/note="MERS3"
/rpc_family="DNA"
/rpc_type=DISPERSED
18847..19585
repeat_region
/note="L1"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
19795..19917
repeat_region
/note="(TA)n"
/rpc_family="Simple_repeat"
/rpc_type=TANDEM
20023..20166
repeat_region
/note="(TA)n"
/rpc_family="Simple_repeat"
/rpc_type=TANDEM
20289..20448
Query Match 86.6%; Score 434; DE 9; Length 340000;
```

```
Best Local Similarity 92.0%; Pred. No. 5.4e-129;
Matches 458; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 4 TCGGCCAACCTCCCAACAGCAGCTAGGTTTCTGTGAGATGGGAGCTGAGAGACAG 63
Db 27679 TTGGCCAACTCTCCCAACAGCAGCTTGGGTTTCTGTGAGAGGGGAGTACGAGAGACAG 27620

QY 64 GACTAGCTGATTTCTAGAGCTGATAGAAATCCCTTAAGCTTACCTGGGAAGTGACAC 123
Db 27619 GACTAGCTGATTTCTAGAGCTGATAGAAATCCCTTAAGCTTACCTGGGAAGTGACAC 27560

QY 124 ATCCACCTTTAAACAGGGGCTTGCACCTTACCTGACACCTGACCAATCAGAGGCTCAC 183
Db 27559 ATCCACCTTTAAACAGGGGCTTGCACCTTACCTGACACCTGACCAATCAGAGGCTCAC 27500

QY 184 TAAATGCTAATTTAGGCAAAAGAGAGTAAAGAAATAGCAATTCATTTCTCTGAG 243
Db 27499 TAAATGCTAATTTAGGCAAAAGAGAGTAAAGAAATAGCAATTCATTTCTCTGAG 27440

QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAATCTTGCAGCCGCAACGCA 303
Db 27439 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAATCTTGCAGCCGCAACGCA 27380

QY 304 ACCCCCTTTGGGTCCTCCCTCTTTGATAGGAGCTGTTTCATGCTATTCTCTAT 363
Db 27379 ACCCCCTTTGGGTCCTCCCTCTTTGATAGGAGCTGTTTCATGCTATTCTCTAT 27320

QY 364 TAAATCTGCAACTGCACTCTTCTGTCATGTTCTTCTTACCGGCTTGAGCTGAGCTTTGCG 423
Db 27319 TAAATCTGCAACTGCACTCTTCTGTCATGTTCTTCTTACCGGCTTGAGCTGAGCTTTGCG 27260

QY 424 TCGCCATCCACCTGCTGTTTGCCTGCGCACCGACAGCCCGCTGACTCCCATCTCT 483
Db 27259 TCGCTTTCACACCTGCTGTTTGCCTGCTGTCGACAGCTGCACTGACTCCCATCTCTCT 27200

QY 484 GGATCATCGAGGTCCTCC 501
Db 27199 GGATCCGCGCAGGCTGCTCC 27182

RESULT 8
AC093531/c 163803 bp DNA linear PRI 16-NOV-2001
LOCUS Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
AC093531
VERSION AC093531.2 GI:16945981
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 163803)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
JOURNAL
Unpublished
2 (bases 1 to 163803)
DOE Joint Genome Institute.
Direct Submission
JOURNAL
Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 163803)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
JOURNAL
Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Nov 16, 2001 this sequence version replaced gi:15383820.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.
Location/Qualifiers
```


source 1.163803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-405L7"

BASE COUNT 54797 a 31372 c 29086 g 48548 t
ORIGIN

Query Match 86.1%; Score 431.2; DB 9; Length 163803;
Best Local Similarity 91.4%; Pred. No. 4e-128;
Matches 457; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 2 TCTGGCCCAACCTCCCAACGACCTTAAAGTTTCTTGAAGTGGGAGCTGAGAC 61
Db 117108 TGTAGCCCAACCTCCCAACGACCTTAAAGTTTCTTGAAGTGGGAGCTGAGAC 117049

QY 62 AGGACTAGCTGATTTCTTAGCTGACCTAAATCCCTAAGCTTAAAGTGGAGTGC 121
Db 117048 AGGACTAGCTGATTTCTTAGCTGACCTAAATCCCTAAGCTTAAAGTGGAGTGC 116989

QY 122 ACATCCACCTTTAAACAAGGGGCTTGCACCTTGAAGCTGACCTGACAGAGAGCTC 181
Db 116988 GCATCCACCTTTAAACAAGGGGCTTGCACCTTGAAGCTGACCTGACAGAGAGCTC 116929

QY 182 ACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCAATCATATTTGCTG 241
Db 116928 ACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCAATCATATTTGCTG 116869

QY 242 AGAGCAGCAGGAGGAGCAATGATCGGATTTAAACCAAGCTTGAAGCGGCAAGG 301
Db 116868 AGAGCAGCAGGAGGAGCAATGATCGGATTTAAACCAAGCTTGAAGCGGCAAGG 116809

QY 302 CAACCCCTTTGGGTCCTCCCTCTTTGATGGAGCTGTTTTCATGCTATTTCACTCT 361
Db 116808 CAACCCCTTTGGGTCCTCCCTCTTTGATGGAGCTGTTTTCATGCTATTTCACTCT 116749

QY 362 ATTAATCTTGCACTGCACTCTTCTGTCATGTTTCTTAAGCTGAGTGAAGCTTTC 421
Db 116748 ATTAATCTTGCACTGCACTCTTCTGTCATGTTTCTTAAGCTGAGTGAAGCTTTC 116689

QY 422 GCTGCCATCCACACCTGCTTTTGGCCGACCGGACCGGCTGACCTCCATCCCT 481
Db 116688 GCTGCCATCCACACCTGCTTTTGGCCGACCGGACCGGCTGACCTCCATCCCT 116629

QY 482 CTGATCATGACGAGGTGCTC 501
Db 116628 TTGATTCACGACGAGGTGCTC 116609

RESULT 9
AF127226 839 bp DNA linear PRI 17-JUN-1999
LOCUS Homo sapiens Human endogenous retrovirus W, 3' long terminal
DEFINITION repeat, partial sequence.
ACCESSION AF127226
VERSION AF127226.1 GI:5081468
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 839)
AUTHORS Ounanian-Paraz, A., Komurian-Pradel, F., Ott, C., Rajorison, A. and Perron, H.
TITLE MSRV-related HERV-W elements in human DNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 839)
AUTHORS Ounanian-Paraz, A., Komurian-Pradel, F., Ott, C., Rajorison, A. and Perron, H.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-1999) UKRI03 CNRS-Biomerieux, Biomerieux, 46, Ailee D' Italie, Lyon 69007, France
FEATURES Location/Qualifiers

source 1.839
/organism="Homo sapiens"
/db_xref="taxon:9606"
LTR <1..>839
/note="3', long terminal repeat; U3, R and U4 regions"

repeat_region 1.839
/note="MSRV-related"
/rpt_family="Human endogenous retrovirus W"
/rpt_type="dispersed"

BASE COUNT 205 a 233 c 192 g 206 t 3 others
ORIGIN

Query Match 85.0%; Score 426; DB 9; Length 839;
Best Local Similarity 90.6%; Pred. No. 9.9e-127;
Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGCCCAACCTCCCAACGACCTTAAAGTTTCTTGAAGTGGGAGCTGAGAC 61
Db 4 TGTAGCCCAACCTCCCAACGACCTTAAAGTTTCTTGAAGTGGGAGCTGAGAC 63

QY 62 AGGACTAGCTGATTTCTTAGCTGACCTAAATCCCTAAGCTTAAAGTGGAGTGC 121
Db 64 AGGACTAGCTGATTTCTTAGCTGACCTAAATCCCTAAGCTTAAAGTGGAGTGC 123

QY 122 ACATCCACCTTTAAACAAGGGGCTTGCACCTTGAAGCTGACCTGACAGAGAGCTC 181
Db 124 GCATCCACCTTTAAACAAGGGGCTTGCACCTTGAAGCTGACCTGACAGAGAGCTC 183

QY 182 ACTAAATGCTAATTAAGCAAGAGGTAAAGAAATAGCAATCATATTTGCTG 241
Db 184 ACTAAATGCTAATTAAGCAAGAGGTAAAGAAATAGCAATCATATTTGCTG 243

QY 242 AGAGCAGCAGGAGGAGCAATGATCGGATTTAAACCAAGCTTGAAGCGGCAAGG 301
Db 244 AGAGCAGCAGGAGGAGCAATGATCGGATTTAAACCAAGCTTGAAGCGGCAAGG 303

QY 302 CAACCCCTTTGGGTCCTCCCTCTTTGATGGAGCTGTTTTCATGCTATTTCACTCT 361
Db 304 CAACCCCTTTGGGTCCTCCCTCTTTGATGGAGCTGTTTTCATGCTATTTCACTCT 363

QY 362 ATTAATCTTGCACTGCACTCTTCTGTCATGTTTCTTAAGCTGAGTGAAGCTTTC 421
Db 364 ATTAATCTTGCACTGCACTCTTCTGTCATGTTTCTTAAGCTGAGTGAAGCTTTC 423

QY 422 GCTGCCATCCACACCTGCTTTTGGCCGACCGGACCGGCTGACCTCCATCCCT 481
Db 424 GCTGCCATCCACACCTGCTTTTGGCCGACCGGACCGGCTGACCTCCATCCCT 483

QY 482 CTGATCATGACGAGGTGCTC 501
Db 484 TTGATTCACGACGAGGTGCTC 503

RESULT 10
AX001030 1329 bp DNA linear PAT 10-MAR-2000
LOCUS AX001030
DEFINITION Sequence 12 from Patent WO9902666.
ACCESSION AX001030
VERSION AX001030.1 GI:7241266
KEYWORDS unclassified.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1329)
AUTHORS Ott, C. and Bedin, F.
TITLE RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
JOURNAL Patent: WO 9902666-A 12 21-JAN-1999;
BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES Location/Qualifiers
source 1.1329
/organism="unclassified"

/db_xref="taxon:32644"
BASE COUNT 346 a 352 c 288 g 340 t 3 others
ORIGIN
Query Match 85.0%; Score 426; DB 6; Length 1329;
Best Local Similarity 90.6%; Pred. No. 1e-126;
Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 2 TCTGGGCACTCCCAAGACGACCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGAC 61
DB TGTACGCCAACCTCCCAACGATCTTGGGTTTCTGTTGAGAGGTTGAGACTGAGAGAC 556
QY 62 AGACCTAGCTGATTTCTAGGCTGACTAAGATCCCTAAGCTTAGCTGGAGAGTGACC 121
DB AGGACTAGCTGATTTCTAGGCTGACTAAGATCCCTAAGCTTAGCTGGAGAGTGACC 616
QY 122 ACATTCACCTTTAAACAGGGGCTTGCACTTAGTGACACTGACCAATCAGAGAGCTC 181
DB GCATTCACCTTTAAACATGGGGCTTGCACTTAGTGACACTGACCAATCAGAGAGCTC 676
QY 182 ACTAAATGCTAATTAAGGCAAGAGAGGTAAAGAAATAGCAATGATCTATGCTG 241
DB ACTAAATGCTAATTAAGGCAAGAGAGGTAAAGCAATGATCTATGCTG 736
QY 242 AGAGCAGCAGAGAGAGCAATGATCGGATATATACCAAGCTCTTGAGCCGCAACGG 301
DB AGAGCAGCAGCGGAGAGCAAGATGGGATATATACCAAGCTCTTGAGCCGCAACAG 796
QY 302 CAACCCCTTTGGGTCCTCTCTTTGATGGAGCTGTTTTCATGCTATTTCACTCT 361
DB CAACCCCTTTGGGTCCTCTCTCTTTGATGGAGCTGTTTTCATGCTATTTCACTCT 856
QY 362 ATTAATCTGCACTGCACTCTTGCTGATGCTATGCTGCTGAGCTGAGCTTTC 421
DB ATTAATCTGCACTGCACTCTTGCTGATGCTATGCTGCTGAGCTGAGCTTTC 916
QY 422 GCTCCGCACTCCACCACTGCTGTTTCCGCGCACCGCAGCCGCTGACTCCCATCTT 481
DB GTTCCGCACTCCACCACTGCTGTTTCCGCGCACCGCAGCCGCTGACTCCCATCTT 976
QY 482 CTGATCATGCAAGGTGTC 501
DB TTGGATCCAGCAGAGTGTTC 996
RESULT 11
AX000970 783 bp DNA linear PAT 10-MAR-2000
LOCUS Sequence 15 from Patent WO9902696.
DEFINITION AX000970
ACCESSION AX000970
VERSION AX000970.1 GI:7241212
KEYWORDS
ORGANISM
SOURCE
unidentified.
unclassified.
REFERENCE
1 (bases 1 to 783)
Beseme, F. and Blond, J.
TITLE
ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL
Patent: WO 9902696-A 15 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source
1..783
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 173 a 213 c 166 g 180 t 51 others
ORIGIN
Query Match 85.0%; Score 425.6; DB 6; Length 783;
Best Local Similarity 94.0%; Pred. No. 1.3e-126;
Matches 422; Conservative 25; Mismatches 1; Indels 1; Gaps 1;
QY 54 TGAGAGACAGACTAGCTGATTTCTAGCTGACTAAGATCCCTAAGCTTAGCTGGGA 113

|||||
DB 1 TGAGAGACAGACTAGCTGATTTCTAGGCTGACTAAGATCCCTAAGCTTAGCTGGGA 60
QY 114 AGTGACCACTCCACTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAATCA 173
DB AGTGACCACTCCACTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAATCA 120
QY 174 GAGAGCTCACTAAATGCTAATTAAGGCAAGACAGAGGTAAAGAAATAGCAATCTCT 233
DB GAGAGCTCACTAAATGCTAATTAAGGCAAGACAGAGGTAAAGAAATAGCAATCTCT 180
QY 234 ATTCCCTGAGAGCAGACAGAGGACAAATGATGGATATTAACCAAGCTTGAGCC 293
DB ATTCCCTGAGAGCAGACAGAGGACAAATGATGGATATTAACCAAGCTTGAGCT 240
QY 294 GGCAAGGCA-ACCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTTTTCATGCTA 352
DB GGCAAGGCAAGCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTTTTCATGCTA 300
QY 353 TTGACTTATTAATCTTGCACTGCACTCTTGCTGCTCAATGTTTTCATGCTGAGC 412
DB TTGACTTATTAATCTTGCACTGCACTCTTGCTGCTCAATGTTTTCATGCTGAGC 360
QY 413 TGAGCTTCCGCTCCGCACTCCACTGCTGTTTGGCGCACCGGACCGCGCTGACT 472
DB TGAGCTTCCGCTCCGCACTCCACTGCTGTTTGGCGCACCGGACCGCGCTGACT 420
QY 473 CCGATCCCTGATGATGAGAGGTGTC 501
DB CCGATCCCTGATGATGAGAGGTGTC 449
RESULT 12
HUA000660 251124 bp DNA linear PRI 15-JUN-2001
LOCUS Homo sapiens T-cell receptor alpha delta locus from bases 501613 to
DEFINITION 752736 (section 3 of 5) of the Complete Nucleotide Sequence.
ACCESSION AE000660 AE000521 U85197
VERSION AE000660.1 GI:2358042
KEYWORDS
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 251124)
Koob, B.F., Rowen, L., Wang, K., Kuo, C.L., Seto, D., Lenzette, J.A.,
Howard, S., Shan, W., Deshpande, P. and Hood, L.
TITLE
The human T-cell receptor TCR α /TCR δ (C alpha/C delta) region:
organization, sequence, and evolution of 97.6 kb of DNA
JOURNAL
Genomics 19 (3), 478-493 (1994)
MEDLINE
94245236
PUBMED
8188290
REMARK
This citation covers from bases 966183-1064019
REFERENCE
AUTHORS
2 (bases 1 to 251124)
Boysen, C., Simon, M.I. and Hood, L.
TITLE
Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with
bacterial artificial chromosome clones
JOURNAL
Genome Res. 7 (4), 330-338 (1997)
MEDLINE
97264339
PUBMED
9101172
REFERENCE
AUTHORS
3 (bases 1 to 251124)
Boysen, C., Inyoub, L., Smith, T.M., Smit, A., Wang, K., Rowen, L. and
Hood, L.
TITLE
T-cell Receptor Alpha Delta Locus Complete Nucleotide Sequence
JOURNAL
Unpublished
REMARK
This citation covers bases 1-983545 and bases 1064020-1071650
AUTHORS
4 (bases 1 to 251124)
Boysen, C., Lee, I., Smith, T.M., Smit, A., Kai, W., Lee, R. and Leroy, H.
TITLE
Direct Submission
JOURNAL
Submitted (20-JUN-1997) Department of Molecular Biotechnology,
University of Washington, Box 357730, Seattle, Washington 98195,
USA
REMARK
Complete nucleotide sequence of the human T-cell receptor alpha


```

repeat_region      13763..13833
                    /rpt_family="MIR2"
repeat_region      13879..13946
                    /rpt_family="MIR"
repeat_region      complement(19382..19677)
                    /rpt_family="AluSg"
repeat_region      21487..21622
                    /rpt_family="MIR"
repeat_region      23034..23706
                    /rpt_family="L1MB3A"
repeat_region      23955..24354
                    /rpt_family="MST1"
repeat_region      complement(24364..24523)
                    /rpt_family="MIR2"
repeat_region      complement(24937..25045)

Query Match      83.0%; Score 415.6; DB 9; Length 251124;
Best Local Similarity 90.5%; Pred. No. 4.9e-123;
Matches 458; Conservative 0; Mismatch 39; Indels 9; Gaps 1;

QY 4 TCGGCAACCTCCCAAGACGACTTAGGTTTCCTTTGAGATGGGGGAGTGAAGACAG 63
Db 247262 TCAGCAACCTCCCAAGACGACTTAGGTTTCCTTTGAGATGGGGGAGTGAAGACAG 247321
QY 64 GACTAGCTGATTTCTCTAGGCTGACTAGATCCCAAGCTTAGCTGGAGTGACCAAC 123
Db 247322 GACTAGCTGATTTCTCTAGGCTGACTAGATCCCAAGCTTAGCTGGAGTGACCAAC 247381
QY 124 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTC/CACCTGACCAATC-----AG 174
Db 247382 TTCCACCTTTAAACAGGGGCTTGCACTTAGCTC/CACCTGACCAATCAGGTAGAAAG 247441
QY 175 AGAGTCTCTAAATGCTAATAGCAAGAGGAGTGAAGAAATAGCAATATCTA 234
Db 247442 AGAGTCTCTAAATGCTAATAGCAAGAGGAGTGAAGAAATAGCAATATCTA 247501
QY 235 TTGCTGAGAGCAGCAGAGGAGCAATGATCGGATTAATTAACCAACTTTGAGCCG 294
Db 247502 TCACCTGAGAGCAGCAGAGGAGCAATGATCGGATTAATTAACCAAGTATTCGAGCTG 247561
QY 295 GCAACGGCAACCCCTTTGGTCCCTCTTTGATGAGGAGCTGTTTTCATGCTATT 354
Db 247562 GCAACGGCAACCCCTTTGGTCCCTCTTTGATGAGGAGCTGTTTTCATGCTATT 247621
QY 365 TCACCTTAATTAATCTTGCACTGCTTCCGTCATGTTTCTTACGGCTGAGCTG 414
Db 247622 TCACCTTAATTAATCTTGCACTGCTTCCGTCATGTTTCTTACGGCTGAGCTG 247681
QY 415 AGCTTTCGCTCGCATCCACCACTGCTGTTGCGCGCAACGAGACCCGCGCTGACTCC 474
Db 247682 AGCTTTCGCTCGCATCCACCACTGCTGTTGCGCGCAACGAGACCCGCGCTGACTTC 247741
QY 475 CATCCCTCTGATCATGCAAGGCTGTC 500
Db 247742 CATCCCTCTGATCATGCAAGGCTGTC 247767

RESULT 13
AC092510/c 187321 bp WNA linear HTG 03-AUG-2002
LOCUS Papio cynocephalus anubis clone RP41-107F15, WORKING DRAFT
DEFINITION SEQUENCE, 5 ordered pieces.
ACCESSION AC092510.2 GI:22094320
VERSION AC092510
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE olive baboon.
ORGANISM Papio cynocephalus anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Papio.
REFERENCE 1 (bases 1 to 187321)
Akhner N., Antonelli A., Ayele K., Beckstrom-Sternberg S.M.,
Benjamin B., Blakesley R.W., Bouffard G.G., Breen K., Brinkley C.,

```

TITLE AUTHORS JOURNAL REFERENCE COMMENT

```

Brooks S., Dietrich N.L., Grant S., Guan X., Gupta J.,
Haghighi P., Hansen N., Ho S.-L., Idol J.R., Karlins E., Laric P.,
Lee-Lin S.-O., Legaspi R., Maduro Q.L., Maduro V.B.,
Margulies E.H., Mastello C., Maskeri B., Mastrian S.D.,
McCluskey J.C., McDowell J., Paguitan C., Pearson R.,
Portnoy M.E., Praead A., Schueler M.G., Stantrop S., Thomas J.W.,
Thomas P.J., Touchman J.W., Tsugeon C., Vogt J.U., Walker M.A.,
Wetherby K.D., Wiggins L., Young A., Zhang L.-H. and Green E.D.
NSC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 187321)
Green E.D.
Direct Submission
Submitted (13-JUL-2001) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 187321)
Green E.D.
Direct Submission
Submitted (03-AUG-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Aug 3, 2002 this sequence version replaced gi:14717323.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@ngri.nih.gov
----- Project Information
Center project name: amf
Center clone name: 107F15

The sequence data in this record represents an 'enhanced'
version of a phase 2 submission. Specifically, the indicated
order and orientation of each sequence contig has been
established using one or more of the following: read-pair
data from individual subclones, overlaps with neighboring
clones, alignment with available reference sequence (e.g.,
human), and/or confirmation by PCR testing. In addition,
the sequence assembly is based on at least 8x average
coverage in Q20 bases and has been reviewed to rule out
gross misassemblies, the low-quality ends of sequence
contigs have been trimmed away, and each base is associated
with a Phrap-derived quality score.
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 185897 bases at least Q40
Consensus quality: 186768 bases at least Q20
Insert size: 165000; agarose-gel
Insert size: 149000; pulse-field-gel
Insert size: 186921; sum-of-contigs
Quality coverage: 12.32x in Q20 bases; agarose-gel
Quality coverage: 13.64x in Q20 bases; pulse-field-gel
Quality coverage: 10.87x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 25945: contig of 25945 bp in length
* 25946 26045: gap of unknown length
* 26046 59962: contig of 33917 bp in length
* 59963 60062: gap of unknown length
* 60063 77104: contig of 17042 bp in length
* 77105 77204: gap of unknown length
* 77205 160478: contig of 83274 bp in length
* 160479 160578: gap of unknown length

```

FEATURES * 160579 187321: contig of 26743 bp in length.
Location/Qualifiers
1. 187321
/organism="Homo sapiens" andbis"
/db_xref="taxon:9555"
/clone="RP41-107F15"
/clone_id="RP41"
1. 25545
/note="assembly_fragment"
clone_end:77
vector_side:left"
1. 24497
/note="clone overlaps with GenBank Accession Number
AC092516 clone RP41-76N20 (center project name ame)"
26046. 59962
/note="assembly_fragment"
60063. 77104
/note="assembly_fragment"
77205. 160478
/note="assembly_fragment"
102673. 187321
/note="clone overlaps with GenBank Accession Number
AC096848 clone RP41-28K15 (center project name aug)"
160579. 187321
/note="assembly_fragment"
clone_end:8P6
vector_side:right"
BASE COUNT 56329 a 38211 c 38592 g 53789 t 400 others
ORIGIN
Query Match 81.8%; Score 409.6; DB 2; Length 187321;
Best Local Similarity 92.3%; Pred. No. 4.1e-121;
Matches 443; Conservative 0; Mismatches 34; Indels 3; Gaps 1;
QY 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGATGCGGAGCTGAGAGACG 63
Db 54177 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGAGAGGAGCACTAAGAGACG 54118
QY 64 GACTAGCTGAGATTCTCTAGGCTGACTAAGAAATCCCTAAGCTAGCTGGGAGAGTACCA 123
Db 54117 GACTAGCTGAGATTCTCTAGGCACTAAGAAATCCCTAAGCTAGCTGGGAGAGTACCA 54058
QY 124 ATCCACCTTTAAACAGCGGGCTTGAACCTTACTCAACCTGACCAATCAGAGAGTCA 183
Db 54057 ATCCACCTTTAAACAGCGGGCTTGAACCTTACTCAACCTGACCAATCAGAGAGTCA 53958
QY 184 TAAATGCTAATTAGGCAAGAGAGAGTAAAGAAATAGCCATCATCTATTGCTGAG 243
Db 53957 TAAATGCTGATTAGGCAAGAGAGAGTAAAGAAATAGCCATCATCTATTGCTGAG 53938
QY 244 AGCAGCAGCAGAGAGAGCAATGATCGGATATTAACCCAGTCTTGAAGCGGCAACGCA 303
Db 53937 AGCAGCAGCAGAGAGAGCAATGATCGGATATTAACCCAGTCTTGAAGCGGCAACGCA 53878
QY 304 ACCGCTTGGGTCCCTCCCTTGTATGAGGAGCTGTGTTTCAATGCTATTCACTCT 363
Db 53877 ACCGCTTGGGTCCCTCCCTTGTATGAGGAGCTGTGTTTCAATGCTATTCACTCT 53818
QY 364 TAAATCTGCAACCTGACTCTTCTGATCAGTCTTCTTAAGGCTGAGAGCTTTCGC 423
Db 53817 TAAATCTGCAACCTGACTCTTCTGATCAGTCTTCTTAAGGCTGAGAGCTTTCGC 53758
QY 424 TCGCATTCACCACTGCTGTTTGGCGGACCGGAGACCGGCGGCTGATCCCATCTCT 483
Db 53757 TCGCATTCACCACTGCTGTTTGGCGGACCGGAGACCGGCGGCTGATCCCATCTCT 53701

RESULT 14
AC092843 83412 bp DNA linear PRI 21-FEB-2002
LOCUS AC092843
DEFINITION Homo sapiens BAC clone RP11-41904 from 2, complete sequence.
ACCESSION AC092843

VERSION AC092843.4 GI:18072221
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 83412)
AUTHORS Sulston, J. B. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 83412)
AUTHORS Cedroni, M. and Haglund, K.
TITLE The sequence of Homo sapiens BAC clone RP11-41904
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 83412)
AUTHORS Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 83412)
AUTHORS Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 83412)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 6, 2002 this sequence version replaced gi:117921256.

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@waterston.wustl.edu

Summary Statistics
Center project name: H_NH0419004

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RP11-41 human BAC library was made from the blood of one male
donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E.,
Tateno, M., Catanesi, J. J., and de Jong, P. J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genome 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.reagen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-512N1, 2000 bp overlap; the clone sequenced to the right is RP11-656O23. Actual start of this clone is at base position 64513 of RP11-512N1; actual end is at base position 83412 of RP11-419O4.

Data from AC068520 was used to finish the clone, AC092843.

```
FEATURES
source
    location/Qualifiers
        1. 83412
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="2"
            /map="2"
            /clone="RP11-419O4"
            /clone_id="RPCT-11"
            251..428
            /rpt_family="MIR"
            434..653
            /rpt_family="MER1_type"
            885..999
            /rpt_family="Alu"
            1382..1416
            /rpt_family="CATATA)n"
            3189..3347
            /rpt_family="MER1_type"
            3745..4097
            /rpt_family="MaLR"
            5089..5311
            /rpt_family="MIR"
            6018..6430
            /rpt_family="ERV1"
            6441..6634
            /note="match to EST BE182495 (NID:98661671)"
            7447..7507
            /rpt_family="ERV1"
            7486..7584
            /rpt_family="TRGA)n"
            7566..7962
            /rpt_family="ERV1"
            8392..8422
            /rpt_family="(TTTG)n"
            8624..8722
            /rpt_family="MER103"
            8837..8984
            /rpt_family="CRI"
            9090..9839
            /rpt_family="ERV1"
            10545..10621
            /rpt_family="GA-rich"
            10557..10583
            /note="similar to Homo sapiens EST BR264057 (NID:99137607)"
            15765..15803
            /rpt_family="ERV1"
            15794..16085
            /rpt_family="ERV1"
            16086..16863
            /rpt_family="ERV1"
            16935..17592
            /rpt_family="CRI"
            17360..18105
            /rpt_family="L2"
            18346..18422
            /rpt_family="MIR"
            18526..18857
            /rpt_family="MaLR"
            18858..19224
            /rpt_family="MaLR"
            18915..18935
            /rpt_family="(A)n"
            19225..19248
            /rpt_family="MaLR"
            19311..19373
            /rpt_family="L1"
```

```
repeat_region 19387..19561 /rpt_family="L1"
repeat_region 20213..20233 /rpt_family="AT-rich"
repeat_region 20763..20932 /rpt_family="MIR"
repeat_region 20980..21353 /rpt_family="MaLR"
repeat_region 21489..21589 /rpt_family="L2"
repeat_region 21668..21783 /rpt_family="MIR"
repeat_region 22245..22318 /rpt_family="MIR"
repeat_region 23790..23942 /rpt_family="MIR"
repeat_region 24196..24218 /rpt_family="L2"
repeat_region 24445..24563 /rpt_family="AT-rich"
repeat_region 24564..24748 /rpt_family="L1"
repeat_region 24749..25103 /rpt_family="MER1_type"
repeat_region 25113..25236 /rpt_family="L1"
repeat_region 28015..28291 /rpt_family="Alu"
repeat_region 31569..33238 /rpt_family="L1"
repeat_region 33626..33996 /rpt_family="MaLR"
repeat_region 34390..34670 /rpt_family="L1"
repeat_region 38082..38216 /rpt_family="r-rich"
repeat_region 40015..40036 /rpt_family="(TTTTA)n"
repeat_region 40025..40800 /rpt_family="L1"
repeat_region 40803..41831 /rpt_family="L1"
repeat_region 41230..41235 /note="similar to Homo sapiens EST AA852771 (NID:92941364)"
repeat_region 41829..42643 /rpt_family="L1"
repeat_region 42646..42832 /rpt_family="L1"
repeat_region 42833..43136 /rpt_family="Alu"

Query Match 81.6%; Score 409; DB 9; Length 83412;
Best Local Similarity 91.4%; Pred. No. 5.8e-121; Indels 3; Gaps 2;
Matches 456; Conservative 0; Mismatches 40;
```

QY 4 TCGGCAACCTCCCAAGACGACCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 16036 TAGGCAACCTCCCAAGACGACCTTAGGTTTCTGTTGAGAGGGGGATGAGAGACAG 16095

QY 64 GACTAGCTGAGATTTCTGAGCTGATAGATCCCTAGAGCTTGGGAGAGTACAC 123
DB 16096 GACTAGCTGAGATTTCTGAGCTGATAGATCCCTAGAGCTTGGGAGAGTACAC 16155

QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 183
DB 16156 ATCCACCTTTAAATACGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 16215

QY 184 TAAATGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCATTTGCTGAG 243
DB 16216 TAAATGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCATTTGCTGAG 16275

QY 244 AGCAGCAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGACCGGCAAGCGCA 303

|||||
Db 16276 AGCAGACGAGGAGGACAAAGATTGGATATATATCCAGGCATTCGAGCTGCACAGCA 16335
Oy 304 ACCCCCTTTGGGTCCCTCCCTTTGATGGAGAGCTGTGTTTCATGCTATTTTCACTTAT 363
Db 16336 ACCCCCTTTGGGTCCCTCCCTTTGATGGAG--CTGTTTTCACCTTATTTTCACTTAT 16393
Oy 364 TAAATCTTGCACTGACCTCTTCTGGTCCATGTTTCTTAACGCTTGAGCTGAGCTTTCGC 423
Db 16394 TAAATCTTGCAACTGACCTCTTCTGGTGCATGTTTGTATACGCTTGAGCTGAACCTTTCAC 16453
Oy 424 TCGCATTCACCACTGCTG--TTTGGCCGACCGCAGACCCGCGCTGACTCCATCCCTC 482
Db 16454 TCGCATTCACCACTGCTGTTTGGCGCGCTGCAGACCACTGCTGACTTCATCTTC 16513
Oy 483 TGGATCATGACGGGTGCC 501
Db 16514 TGGATTCAGCAGGGTGCC 16532

RESULT 15
AL139038 140756 bp DNA linear PRI 15-JUL-2001
LOCUS AL139038
DEFINITION Human DNA sequence from clone RP11-456B18 on chromosome 13,
complete sequence.
ACCESSION AL139038
VERSION AL139038.18 GI:14800148
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 140756)
Dunn,M.
Direct Submission
Submitted (14-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Jul 17, 2001 this sequence version replaced gi:13751261.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-456B18 is from the library RPC1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-456B18. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-456B18 is at 140756 in this
sequence. The true left end of clone RP11-78L16 is at 68441 in this
sequence. The true right end of clone RP11-108H9 is at 100 in this
sequence.
Location/Qualifiers

source
1. .140756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-456B18"
/clone_id="RPC1-11.2"
1830. .2326
/note="LOR1a repeat: matches 1. .497 of consensus"
3932. .3965
/note="17 copies 2 mer gt 94% conserved"
4632. .5067
/note="LOR1b repeat: matches 1. .461 of consensus"
5733. .5893
/note="L1ME3 repeat: matches 6002. .6146 of consensus"
6000. .6083
/note="MER34 repeat: matches 454. .539 of consensus"
6108. .6171
/note="HERV23 repeat: matches 993. .1056 of consensus"
7143. .7840
/note="LTR8 repeat: matches 1. .691 of consensus"
8059. .8292
/note="LTR16a repeat: matches 203. .442 of consensus"
8491. .8686
/note="LTR29 repeat: matches 6. .203 of consensus"
8725. .8926
/note="MER4D repeat: matches 483. .677 of consensus"
8992. .9049
/note="MER4D repeat: matches 410. .465 of consensus"
9062. .9281
/note="110 copies 2 mer tt 59% conserved"
9398. .9530
/note="MER41C repeat: matches 1. .131 of consensus"
9623. .9732
/note="LTR1 repeat: matches 675. .785 of consensus"
9770. .10333
/note="LTR2D repeat: matches 1. .486 of consensus"
10400. .10545
/note="LTR29 repeat: matches 2. .151 of consensus"
10549. .10935
/note="MER67D repeat: matches 12. .391 of consensus"
11249. .11332
/note="MIR repeat: matches 69. .144 of consensus"
11923. .12159
/note="MER4D repeat: matches 1. .234 of consensus"
12164. .12767
/note="MER4D repeat: matches 396. .973 of consensus"
12852. .13221
/note="THR1B repeat: matches 1. .364 of consensus"
13222. .14798
/note="THR1B-INTERNAL repeat: matches 1. .1580 of
consensus"
14789. .15173
/note="THR1B repeat: matches 1. .364 of consensus"
16144. .16346
/note="MER54B repeat: matches 585. .789 of consensus"
16339. .16503
/note="MER54B repeat: matches 112. .278 of consensus"
16504. .16571
/note="MER66A repeat: matches 1. .69 of consensus"
16570. .16629
/note="MER41B repeat: matches 283. .341 of consensus"
16599. .16902
/note="MER66A repeat: matches 163. .478 of consensus"
16903. .17014
/note="MER54B repeat: matches 3. .115 of consensus"
17195. .17491
/note="AluY repeat: matches 1. .297 of consensus"
18120. .19291
/note="L1p repeat: matches 1726. .2897 of consensus"
19460. .19584
/note="25 copies 5 mer attt 62% conserved"
19686. .19745
/note="12 copies 5 mer tttaa 71% conserved"

```
repeat_region 20453. .20914
/note="L1MD2 repeat: matches 5745. .6199 of consensus"
repeat_region 20913. .21222
/note="L1MD2 repeat: matches 5277. .5594 of consensus"
repeat_region 21308. .21570
/note="L1R9 repeat: matches 357. .625 of consensus"
repeat_region 21645. .21984
/note="L1R19B repeat: matches 55. .393 of consensus"
repeat_region 22081. .22583
/note="L1MD2 repeat: matches 4726. .5242 of consensus"
repeat_region 23498. .23784
/note="L1ubd repeat: matches 1. .305 of consensus"
repeat_region 23873. .24169
/note="L1usx repeat: matches 1. .295 of consensus"
repeat_region 25044. .25238
/note="MIR repeat: matches 31. .238 of consensus"
repeat_region 25404. .25696
/note="L1usx repeat: matches 20. .311 of consensus"
repeat_region 25832. .26218
/note="MSTC repeat: matches 1. .405 of consensus"
repeat_region 26221. .27865
/note="MSTC-internal repeat: matches 2. .1651 of consensus"
repeat_region 27907. .28659
/note="L1P82 repeat: matches 5400. .6155 of consensus"
repeat_region 28710. .29043
/note="MSTD repeat: matches 30. .394 of consensus"
repeat_region 29634. .29697
/note="L1P82 repeat: matches 2 mer t a 79% conserved"
repeat_region 29729. .29752
/note="L1P82 repeat: matches 2 mer t g 95% conserved"
repeat_region 30688. .30874
/note="MER61A repeat: matches 160. .354 of consensus"
repeat_region 31648. .31785
/note="L1M4 repeat: matches 6409. .6540 of consensus"
repeat_region 31813. .32243
/note="L1M4 repeat: matches 6629. .7089 of consensus"
repeat_region 32449. .32821
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 32824. .34464
/note="THE1C-internal repeat: matches 3. .1651 of consensus"
repeat_region 34465. .34825
/note="THE1C repeat: matches 1. .360 of consensus"
repeat_region 34804. .35004
/note="MER67C repeat: matches 202. .403 of consensus"
repeat_region 39792. .39868
/note="L1P82 repeat: matches 6076. .6152 of consensus"
repeat_region 40953. .41258
/note="L1uv repeat: matches 1. .308 of consensus"
repeat_region 41279. .41308
/note="L15 copies 2 mer aa 86% conserved"
repeat_region 42744. .42933
/note="L1M8 repeat: matches 6064. .6256 of consensus"
repeat_region 42973. .43275
/note="L1usx repeat: matches 1. .303 of consensus"
repeat_region 43560. .43911
/note="match: STS: Em:HSPD07E5"
misc_feature 45196. .45584
/note="M12B repeat: matches 2. .399 of consensus"
repeat_region 45594. .45767
/note="87 copies 2 mer ta 77% conserved"
repeat_region 45800. .45851
/note="M12B repeat: matches 394. .448 of consensus"
repeat_region 48118. .48487
/note="THE1C repeat: matches 1. .370 of consensus"
repeat_region 48708. .48806
/note="L1ubd/FRAM repeat: matches 197. .297 of consensus"
repeat_region 49853. .50271
/note="L1MEC repeat: matches 2274. .2369 of consensus"
repeat_region 50637. .50927
/note="L1ME repeat: matches 974. .1253 of consensus"
repeat_region 52250. .52349
/note="MIR repeat: matches 131. .228 of consensus"
```

```
repeat_region 52696. .52996
/note="L1usx repeat: matches 1. .293 of consensus"
repeat_region 54927. .55010
Query Match 81.3% Score 407.2; DB 9; Length 140756;
Best Local Similarity 91.9%; Pred. No. 2.4e-120;
Matches 441; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 4 TCGGCAACTGCCCAAGCAGCACTTAGGTTTCCTGTTGAGATGGGGAGCTGAGACAG 63
DB 69419 TTGGCAACTGCCCAAGCAGCACTTAGGTTTCCTGTTGAGATGGGGAGCTGAGACAG 69478
QY 64 GACTAGCTGATTTCTTGGCTGACTAGTAAGATTCCTAGCCTAGTGGGAAGTACCA 123
DB 69479 GACTAGCTGATTTCTTGGCTGACTAGTAAGATTCCTAGCCTAGTGGGAAGTACCA 69538
QY 124 ATCCACTTTAAACGCGGGGCTGCACTTAGCTGACACCTGACCAATCAGAGACTGAC 183
DB 69539 ATCCACTTTAAACGCGGGGCTGCACTTAGCTGACACCTGACCAATCAGAGACTGAC 69598
QY 184 TAAATGCTAATTAAGCAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 243
DB 69599 TAAATGCTAATTAAGCAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAG 69658
QY 244 AGCAGCAGAGAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 303
DB 69659 AGCAGCAGAGAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 69718
QY 304 ACCGCTTTGGTGGTCCCTCCCTGTTGATGGAGCTCTGTTGATGATTAATTAATTA 363
DB 69719 ACCGCTTTGGTGGTCCCTCCCTGTTGATGGAGCTCTGTTGATGATTAATTAATTA 69778
QY 364 TAAATCTGCAACTGCACTCTTGTGTCATGATTTCTTACGCTTGAAGCTGAGC-TTTCG 422
DB 69779 TAAATCTGCAACTGCACTCTTGTGTCATGATTTCTTACGCTTGAAGCTGAGC-TTTCG 69838
QY 423 CTCGCCATCCACCACTGCTGTTGTCGCCGACCGACGACCGCGCTGACTCCATCCCTC 482
DB 69839 CTCGCCATCCACCACTGCTGTTGTCGCCGACCGACGACCGCGCTGACTCCATCCCTC 69898
```

Search completed: April 17, 2003, 06:48:14
Job time : 2133.47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: April 17, 2003, 02:30:00 ; Search time 180.787 Seconds
(without alignments)
6240.778 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501
Sequence: 1 gtcgcgcacaccccccacccca.....ctgcatcatgcaggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq.101002.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	10499	ABN97929	Human retroviral s
2	498	99.4	56093	ABL61744	Colon adenocarcino
3	461.4	92.1	7466	AA68626	DNA encoding novel
4	450	89.8	711	ABN97947	Human retroviral s
5	426	85.0	1329	AAV43219	Multiple sclerosis
6	426	85.0	1329	AA28704	Clone 5M6 from MSR
7	425.6	85.0	783	AA25669	Human endogenous r
8	396.2	79.1	711	ABN97946	Human retroviral s
9	396.2	79.1	1393	AA531002	Human diagnostic a

10	377.8	75.4	893	23	AA65964	DNA encoding novel
11	377	75.2	2781	22	AA65530	Nucleotide sequenc
12	375.4	74.9	2946	20	AA67526	Human secreted pro
13	375.4	74.9	2946	21	AA259468	Human secreted pro
14	374.4	74.7	2930	24	AA24195	Human syncytin cDN
15	373.4	74.5	7582	20	AA25565	Complete human end
16	373.4	74.5	7582	21	AA459215	Human endogenous r
17	372.6	74.4	849	22	AA459215	Human diagnostic a
18	372.2	74.3	3372	20	AA25663	Human endogenous r
19	372.2	74.3	3372	21	AA459213	Partial pol gene a
20	371.2	74.1	1136	20	AA25660	Human endogenous r
21	371.2	74.1	1136	21	AA459210	Human endogenous r
22	371.2	74.1	1136	22	AA459210	3' pol gene and 3'
23	371.2	74.1	5154	23	AA459210	DNA encoding novel
24	371.2	74.1	8279	23	AA67609	DNA encoding novel
25	371.2	74.1	8294	23	AA67609	DNA encoding novel
26	370.6	74.0	2782	20	AA25661	DNA encoding novel
27	370.6	74.0	2782	21	AA459211	Human endogenous r
28	370.6	74.0	2782	22	AA459211	5' non coding, 3'
29	350.8	70.0	2300	23	AA459211	HERV-W envelope pr
30	347.4	69.3	1165	23	AA459211	DNA encoding novel
31	346	69.1	2942	23	AA677313	DNA encoding novel
32	345.8	69.0	2629	22	AA677313	DNA encoding novel
33	341.8	68.2	635	19	AA43215	Human immune/haema
34	341.8	68.2	635	20	AA43215	Multiple sclerosis
35	341.8	68.2	2030	21	AA43215	Nucleotide sequenc
36	325.6	65.0	808	23	AA68392	DNA encoding novel
37	321.8	64.2	1478	23	AA68392	DNA encoding novel
38	320.8	64.0	1243	23	AA684189	DNA encoding novel
39	312.6	62.4	448	23	AA687367	DNA encoding novel
40	312.6	62.4	583	23	AA688391	DNA encoding novel
41	311.4	62.2	410	24	AA694040	Gene #538 used to
42	311.4	62.2	410	24	AA694040	Human benign prost
43	308.2	61.5	17758	22	AA64210	Human immune/haema
44	303.4	60.6	439	22	AA63993	Human polynucleoti
45	303.4	60.6	439	22	AA631671	Genomic sequence #

ALIGNMENTS

RESULT 1
ID ABN97929 standard; DNA; 10499 BP.
AC ABN97929;
DT 01-AUG-2002 (first entry)
DE Human retroviral sequence HERV-7g.
KW Autoimmune disease; HERV-7g; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
OS Human retrovirus.
PN WO9967395-A1.
PD 29-DEC-1999.
PF 23-JUN-1999; 99WO-FR01513.
PR 23-JUN-1998; 98FR-0007920.
PA (INRM) INSERM NAT SANTE & RECH MEDICALE.
PI Alliel PM, Perin J, Rieger F;
DR WPI; 2000-160587/14.
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7g,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -

XX Claim 3; Fig 1; 225pp; French.
XX
PS The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q,
CC regulatory elements associated with HERV-7q may alter expression of other
CC genes (seven remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;

Query Match 100.0%; Score 501; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 8,2e-161;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTGGCCACCTTCCCAACAGCACTTAGTTTCTCTTTAGATGGGAGCTGAGAGA 60
DB 9500 GTCTGGCCACCTTCCCAACAGCACTTAGTTTCTCTTTAGATGGGAGCTGAGAGA 9559
QY 61 CAGGACTAGTGGATTCTTGAAGCTGAGTAAAGCCCAAGCTTGGAGAGGTAC 120
DB 9560 CAGGACTAGTGGATTCTTGAAGCTGAGTAAAGCCCAAGCTTGGAGAGGTAC 9619
QY 121 CACATCCACCTTTAAACAGCGGGCTTCACTTAGCTCACCTGACCAATGAGAGCT 180
DB 9620 CACATCCACCTTTAAACAGCGGGCTTCACTTAGCTCACCTGACCAATGAGAGCT 9679
QY 181 CACTAAATGCTTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTT 240
DB 9680 CACTAAATGCTTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTT 9739
QY 241 GAGAGACAGCAGGAGGAGGAGCAGTATGAGATTAACCAAGTCTTGGAGCGGCAACG 300
DB 9740 GAGAGACAGCAGGAGGAGGAGCAGTATGAGATTAACCAAGTCTTGGAGCGGCAACG 9799
QY 301 GCAACCCCTTTGGTCCCTCCCTTTTATGAGGAGCTCTGTTTTCATCTATTTCATCTC 360
DB 9800 GCAACCCCTTTGGTCCCTCCCTTTTATGAGGAGCTCTGTTTTCATCTATTTCATCTC 9859
QY 361 TATTAAATCTTGCAACTGCACTCTTGTGCAATGTTCTTAAGCGCTTGAAGCTTT 420
DB 9860 TATTAAATCTTGCAACTGCACTCTTGTGCAATGTTCTTAAGCGCTTGAAGCTTT 9919
QY 421 CGCTGGCCATCCACCACTGCTTTGGCGGCAACCGCGCGCTGACCTCCATCCG 480
DB 9920 CGCTGGCCATCCACCACTGCTTTGGCGGCAACCGCGCGCTGACCTCCATCCG 9979
QY 481 TCTGATCATGAGGGGTGTC 501
DB 9980 TCTGATCATGAGGGGTGTC 10000

KW gene; ds.
XX
OS Homo sapiens.
XX WO200194629-A2.
XX
PN 13-DEC-2001.
XX
PD 30-MAY-2001; 2001WO-US10838.
PF
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-235133P.
XX 18-SEP-2000; 2000US-235617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 20-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 25-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235863P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.
XX 28-SEP-2000; 2000US-236109P.
XX 28-SEP-2000; 2000US-236111P.
XX 29-SEP-2000; 2000US-236842P.
XX 29-SEP-2000; 2000US-236891P.
XX 02-OCT-2000; 2000US-237172P.
XX 02-OCT-2000; 2000US-237173P.
XX 02-OCT-2000; 2000US-237278P.
XX 02-OCT-2000; 2000US-237294P.
XX 02-OCT-2000; 2000US-237295P.
XX 02-OCT-2000; 2000US-237316P.
XX 02-OCT-2000; 2000US-237425P.
XX 03-OCT-2000; 2000US-237598P.
XX 03-OCT-2000; 2000US-237604P.
XX 03-OCT-2000; 2000US-237606P.
XX 03-OCT-2000; 2000US-237608P.
XX 01-NOV-2000; 2000US-244867P.
XX 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI, 2002-188264/24.
XX
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set
XX
XX Claim 1; SEQ ID 81; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (i)
XX comprises a sequence (S) selected from 8447 sequences (given in AB161664
XX to AB170110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (i) has cytostatic

CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.

XX
XX Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16861 T; 0 other;
SQ

Query Match 99.4%; Score 498; DB 24; Length 56093;
Best Local Similarity 100.0%; Pred. No. 2,2e-159;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGTGGGGAGCTGAGAGACAG 63
DB 37504 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGTGGGGAGCTGAGAGACAG 37563
QY 64 GACTAGCTGATTTCTTGAAGCTGATTAAGATCCCTAAGCTGAGTGGGAGTACAC 123
DB 37564 GACTAGCTGATTTCTTGAAGCTGATTAAGATCCCTAAGCTGAGTGGGAGTACAC 37623
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAACCTGACCAATCAGAGAGCTCAC 183
DB 37624 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAACCTGACCAATCAGAGAGCTCAC 37683
QY 184 TAAATGCTAATTAGGCAAGAAGAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 243
DB 37684 TAAATGCTAATTAGGCAAGAAGAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 37743
QY 244 AGCAGAGAGAGGAGCAATGATCGGGATATTAACCAAGTTTGGAGCCGGCAACGGCA 303
DB 37744 AGCAGAGAGAGGAGCAATGATCGGGATATTAACCAAGTTTGGAGCCGGCAACGGCA 37803
QY 304 ACCCCCTTTGGGTCCTCCCTGTTGATGGAGCTGTGTTTCATGCTATTTCATCTAT 363
DB 37804 ACCCCCTTTGGGTCCTCCCTGTTGATGGAGCTGTGTTTCATGCTATTTCATCTAT 37863
QY 364 TAAATCTTGCAACTGCACTTCTGTGATGATGTTTCTTAACGGCTTGAGTGAAGCTTGC 423
DB 37864 TAAATCTTGCAACTGCACTTCTGTGATGATGTTTCTTAACGGCTTGAGTGAAGCTTGC 37923
QY 424 TCGCATTCACCACTGCTGTTTGGCCGCAACCGCAACCGCGCTGACTCCATCCCTCT 483
DB 37924 TCGCATTCACCACTGCTGTTTGGCCGCAACCGCAACCGCGCTGACTCCATCCCTCT 37983
QY 484 GGATCATGAGGGGTGCC 501
DB 37984 GGATCATGAGGGGTGCC 38001

RESULT 3
AAS68626
ID AAS68626 standard; cDNA, 7466 BP.
XX
XX AAS68626;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #4430.
DB
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200175067-R2.
EN
XX
XX 11-OCT-2001.
PD
XX

PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PI
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR
XX P-PSDB; ABG04439.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX
XX Claim 1; SEQ ID No 4430; 103pp; English.

XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridization probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WRO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ
SQ Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;
Query Match 92.1%; Score 461.4; DB 23; Length 7466;
Best Local Similarity 96.6%; Pred. No. 2.8e-147;
Matches 482; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGTGGGGAGCTGAGAGACAG 63
DB 3668 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGTGGGGAGCTGAGAGACAG 37227
QY 64 GACTAGCTGATTTCTTGAAGCTGATTAAGATCCCTAAGCTGAGTGGGAGTACAC 123
DB 3728 GACTAGCTGATTTCTTGAAGCTGATTAAGATCCCTAAGCTGAGTGGGAGTACAC 3787
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAACCTGACCAATCAGAGAGCTCAC 183
DB 3788 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAACCTGACCAATCAGAGAGCTCAC 3847
QY 184 TAAATGCTAATTAGGCAAGAAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 243
DB 3848 TAAATGCTAATTAGGCAAGAAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 3907
QY 244 AGCAGAGAGAGGAGCAATGATCGGGATATTAACCAAGTCTTGAGCCGGCAACGGCA 303
DB 3908 AGCAGAGAGAGGAGCAATGATCGGGATATTAACCAAGTCTTGAGCCGGCAACGGCA 3967
QY 304 -ACCCCTTTGGGTCCTCCCTGTTGATGGAGCTGTGTTTCATGCTATTTCCTGTA 362
DB 3968 GCCCCTTTGGGTCCTCCCTGTTGATGGAGCTGTGTTTCATGCTATTTCCTGTA 4027
QY 363 TTAATCTTGCAACTGCACTTCTGTGATCAATGTTTCTTACGGCTTGAGCTGAGCTTTCG 422

```

Db 4028 TTTAACTTGGCACTGCTTCTGATCCATGTTCTTACGGCTGAGCTTG 4087
Qy 423 CTGGCATTCACACCTGCTGTTTGGCGGACCGGAGACCGCGCGGTGACTCCATCCCTC 482
Db 4088 CTCACCGTCACACCTGCTGTTTGGCGGACCGGAGACCTGCGCTGACTCCATCCCTC 4147
Qy 483 TGGATCATGACGAGGTGTC 501
Db 4148 TGGATCTGACGAGGTGTC 4166

RESULT 4
ABN97947
ID ABN97947 standard; DNA; 711 BP.
AC ABN97947;
XX 01-AUG-2002 (first entry)
DE Human retroviral sequence R1F.
XX
XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KM multiple sclerosis; ds.
XX Human retrovirus.
OS
XX MO9967395-A1.
XX 29-DEC-1999.
XX 23-JUN-1999; 99WO-FR01513.
XX 23-JUN-1998; 98FR-0007920.
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX Alliel PM, Perin J, Rieger F;
XX WPI; 2000-160587/14.
XX
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -
XX
XX Claim 3; Fig 3; 225pp; French.
XX
XX The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HIV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer.
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
XX Sequence 711 BP; 171 A; 209 C; 160 G; 171 T; 0 other;
XX
XX Query Match 89.8%; Score 450; D3 21; Length 711;
XX Best Local Similarity 100.0%; Pred. No. 6.1e-144; Indels 0; Gaps 0;
XX Matches 450; Conservative 0; Mismatches 0;

```

```

Qy 172 CAGAGAGCTCACTAAATGCTAATTAGCAAGAAGAGGTTAAAGAAATAGCCATCAT 231
Db 121 CAGAGAGCTCACTAAATGCTAATTAGCAAGAAGAGGTTAAAGAAATAGCCATCAT 180
Qy 232 CTATTGCTTGAAGACACAGCAGAGGAGCAATGATCGGATATTAACCAAGCTCTTGAG 291
Db 181 CTATTGCTTGAAGACACAGCAGAGGAGCAATGATCGGATATTAACCAAGCTCTTGAG 240
Qy 292 CCGGCAACGGACACCCCTTTGGGTCCTCCCTTTGATGGAGCTGTGTTTCATGCT 351
Db 241 CCGGCAACGGACACCCCTTTGGGTCCTCCCTTTGATGGAGCTGTGTTTCATGCT 300
Qy 352 ATTTCACCTCTAATAATCTTCAACTGACACTCTCTGCTCCATGTTTCTTAAGGCTTGAG 411
Db 301 ATTTCACCTCTAATAATCTTCAACTGACACTCTCTGCTCCATGTTTCTTAAGGCTTGAG 360
Qy 412 CTGAGCTTTCGCTGCGCATCCACCACTGCTGTTGGCCGACCGGACCGGCGCTGAC 471
Db 361 CTGAGCTTTCGCTGCGCATCCACCACTGCTGTTGGCCGACCGGACCGGCGCTGAC 420
Qy 472 TCCCATCCCTCGATCATGAGGCTGTC 501
Db 421 TCCCATCCCTCGATCATGAGGCTGTC 450

RESULT 5
AAV43219
ID AAV43219 standard; CDNA; 1329 BP.
XX
XX AAV43219;
XX
XX 29-DEC-1998 (first entry)
XX
XX Multiple sclerosis associated retrovirus fragment 6.
XX
XX Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
KM gag gene; env gene; rheumatoid arthritis-associated virus; ss.
XX
XX Multiple sclerosis associated retrovirus.
XX
XX Key Location/Qualifiers
XX CDS 2..490
XX /tag= A
XX /product= "Encodes protein AAV71069"
XX /transl_except= (pos:77-79, appears to code for a
XX stop codon)
XX /transl_except= (pos:125-127, appears to code for a
XX stop codon)
XX /transl_except= (pos:137-139, appears to code for a
XX stop codon)
XX
XX WO9823755-A1.
XX
XX 04-JUN-1998.
XX
XX 26-NOV-1997; 97WO-IB01482.
XX
XX 26-NOV-1996; 96US-0756429.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F;
XX Mandrand B, Paranhos-Baccala G, Perron H;
XX WPI; 1998-322732/28.
XX P-PSDB; AAV71069.
XX
XX New nucleic acid from retroviruses - useful for diagnosis,
XX prevention and treatment of, e.g. multiple sclerosis
XX
XX Disclosure; Pages 187-188; 286pp; English.
XX

```

CC The present sequence represents a multiple sclerosis (MS) associated
CC retrovirus (MSRV) genomic fragment used in the method of the
CC invention. The invention provides complete or partial genomic
CC sequences of the MSRV-1 pol gene, gag gene and env gene, and
CC polypeptides encoded by these genes. The invention also provides
CC antibodies raised against the polypeptides. The genomic sequences,
CC polypeptides and antibodies are also claimed useful for diagnosing
CC infection by MS and rheumatoid arthritis-associated viruses, and also
CC for prevention and treatment of infection with these viruses.

XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 85.0%; Score 426; DB 19; Length 1329;

Best Local Similarity 90.6%; Pred. No. 1.6e-135; Mismatches 453; Conservative 0; Matches 47; Indels 0; Gaps 0;

QY 2 TCTCGGCCAACCTCCCAAGCACTTAGGTTTCTGTTGAGATGGGGAGTGAAGAC 61
DB TGTGAGCCAACTCCCAAGCACTTAGGTTTCTGTTGAGATGGGGAGTGAAGAC 556
QY 62 AGAAGTACTGATTTCTTAGGCTGACTAAGATCCCTTAAGCTTGGGAAGTGAC 121
DB 557 AGGACTACTGATTTCTTAGGCTGACTAAGATCCCTTAAGCTTGGGAAGTGAC 616
QY 122 ACATCCACCTTTAAACAGGGGCTTGAACCTTAGCTCAACCTGACCAATCAGAGCTC 181
DB 617 GCATCATCTTTAAACATGAGGCTTGAACCTTAGCTCAACCTGACCAATCAGAGCTC 676
QY 182 ACTAAATGCTTAATAGCAAGCAAGAGGTAAAGAAATAGCCATCATCTATTGCTCTG 241
DB 677 ACTAAATGCTTAATAGCAAGCAAGAGGTAAAGAAATAGCCATCATCTATTGCTCTG 736
QY 242 AGAGCAGCAGGAGGAGCAATGATCGGATATTAACCAATGCTTGGACCGGCAACGG 301
DB 737 AGAGCAGCAGGAGGAGCAATGATCGGATATTAACCAATGCTTGGACCGGCAACGG 796
QY 302 CAACCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTTTCAATGCTATTGACTCT 361
DB 797 CAACCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTTTCAATGCTATTGACTCT 856
QY 362 ATTAATCTTGCAATGCACTTTCTGCTGCAATGTTTCTTAGGCTTGAAGTGAAGCTTTC 421
DB 857 ATTAATCTTGCAATGCACTTTCTGCTGCAATGTTTCTTAGGCTTGAAGTGAAGCTTTC 916
QY 422 GCTGGCATTCCACCACTGCTGTTTGGCGCCACCGAGACCGCGCTGACTCCATCCCT 481
DB 917 GTTCCGATCCACCACTGCTGTTTGGCGCCACCGAGACCGCGCTGACTCCATCCCT 976
QY 482 CTGATCATGACGAGGTGTC 501
DB 977 TTGATCCAGCAGAGTGTCC 996

RESULT 6
AAK29704
ID AAK29704 standard; DNA; 1329 BP.

XX AAK29704;

XX AC 08-JUN-1999 (first entry)
XX DT
XX DE Clone 5M6 from MSRV-1.

XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
XX Rheumatoid polyarthritits; ss.

XX Multiple sclerosis related virus type 1.

XX FR2765588-A1.

XX 08-JAN-1999.

XX 07-JUL-1997; 97FR-0008816.

XX 07-JUL-1997; 97FR-0008816.

XX (INMR) BIO MERIEUX.

XX WPI; 1999-098275/09.

XX P-PSDB; AAMW9554.

XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
XX multiple sclerosis or rheumatoid polyarthritits

XX Claim 1; Page 39-40; 83pp; French.

XX This sequence represents clone 5M6 from a novel multiple sclerosis
XX related virus type 1 (MSRV1). The sequence can be used in diagnostic,
XX prophylactic or therapeutic compositions to inhibit expression of a
XX multiple sclerosis related virus and/or virus associated with
XX rheumatoid polyarthritits.

XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 85.0%; Score 426; DB 20; Length 1329;

Best Local Similarity 90.6%; Pred. No. 1.6e-135; Mismatches 453; Conservative 0; Matches 47; Indels 0; Gaps 0;

QY 2 TCTCGGCCAACCTCCCAAGCACTTAGGTTTCTGTTGAGATGGGGAGTGAAGAC 61
DB 497 TGTGAGCCAACTCCCAAGCACTTAGGTTTCTGTTGAGATGGGGAGTGAAGAC 556
QY 62 AGAAGTACTGATTTCTTAGGCTGACTAAGATCCCTTAAGCTTGGGAAGTGAC 121
DB 557 AGGACTACTGATTTCTTAGGCTGACTAAGATCCCTTAAGCTTGGGAAGTGAC 616
QY 122 ACATCCACCTTTAAACAGGGGCTTGAACCTTAGCTCAACCTGACCAATCAGAGCTC 181
DB 617 GCATCATCTTTAAACATGAGGCTTGAACCTTAGCTCAACCTGACCAATCAGAGCTC 676
QY 182 ACTAAATGCTTAATAGCAAGCAAGAGGTAAAGAAATAGCCATCATCTATTGCTCTG 241
DB 677 ACTAAATGCTTAATAGCAAGCAAGAGGTAAAGAAATAGCCATCATCTATTGCTCTG 736
QY 242 AGAGCAGCAGGAGGAGCAATGATCGGATATTAACCAATGCTTGGACCGGCAACGG 301
DB 737 AGAGCAGCAGGAGGAGCAATGATCGGATATTAACCAATGCTTGGACCGGCAACGG 796
QY 302 CAACCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTTTCAATGCTATTGACTCT 361
DB 797 CAACCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTTTCAATGCTATTGACTCT 856
QY 362 ATTAATCTTGCAATGCACTTTCTGCTGCAATGTTTCTTAGGCTTGAAGTGAAGCTTTC 421
DB 857 ATTAATCTTGCAATGCACTTTCTGCTGCAATGTTTCTTAGGCTTGAAGTGAAGCTTTC 916
QY 422 GCTGGCATTCCACCACTGCTGTTTGGCGCCACCGAGACCGCGCTGACTCCATCCCT 481
DB 917 GTTCCGATCCACCACTGCTGTTTGGCGCCACCGAGACCGCGCTGACTCCATCCCT 976
QY 482 CTGATCATGACGAGGTGTC 501
DB 977 TTGATCCAGCAGAGTGTCC 996

RESULT 7
AAK25669
ID AAK25669 standard; cDNA to mRNA; 783 BP.

XX AAK25669;

XX 21-MAY-1999 (first entry)

XX Human endogenous retrovirus W long terminal repeat region.

XX Clone; human endogenous retrovirus; genome; autoimmune disease;

KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
 OS Human endogenous retrovirus.

XX MO9902696-A1.

XX 21-JAN-1999.

PF 06-JUL-1998; 98MO-FR01442.

XX 07-JUL-1997; 97FR-0008815.

PA (INMR) BIO MERIEUX.

PI Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;

DR WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -
 PT expressed exclusively in placenta and useful in diagnosis and
 PT therapy of autoimmune disease, and abnormal or failed pregnancy

XX Claim 1; Page 83; 106pp; French.

XX This sequence represents the long terminal repeat (LTR) region of the
 CC human endogenous retrovirus (HERV) W genome. The nucleic acids, their
 CC fragments or peptides encoded by them are markers of autoimmune disease
 CC (e.g. multiple sclerosis, Rheumatoid polyarthritis, disseminated lupus
 CC erythematosus, insulin-dependent diabetes and related pathologies) and
 CC of abnormal or unsuccessful pregnancy and can be used as chromosomal
 CC makers for susceptibility to these conditions, or proximity markers
 CC of genes associated with this susceptibility.

XX Sequence 783 BP; 173 A; 213 C; 166 G; 180 T; 51 other;

Query Match 85.0%; Score 425.6; DB 20; Length 783;
 Best Local Similarity 94.0%; Pred. No. 1.6e-135;
 Matches 422; Conservative 25; Mismatches 1; Indels 1; Gaps 1;

QY 54 TGAGAGACAGAGCTAGCTGATTTCTTAGCTGACTAAGATCCCTAAGCTTACGCTGGA 113

DB 1 TGAGAGACAGAGCTAGCTGATTTCTTAGCTGACTAAGATCCCTAAGCTTACGCTGGA 60

QY 114 AGGTACCAACATCCACCTTAAACAGGGGCTTGCACTTAGCTCAGCTGACCATCA 173

DB 61 AGGTACCAACATCCACCTTAAACAGGGGCTTGCACTTAGCTCAGCTGACCATCA 120

QY 174 GAGAGCTCACTAAATGCTAATTAGCAAGAGAGTAAAGAAATAGCCATCATCT 233

DB 121 GAGAGCTCACTAAATGCTAATTAGCAAGAGAGTAAAGAAATAGCCATCATCT 180

QY 234 ATTGCTGAGAGACAGAGGAGGACATATGCGATATATAACCAAGCTTGGAGCC 293

DB 181 ATTGCTGAGAGACAGAGGAGGACATATGCGATATATAACCAAGCTTGGAGCC 240

QY 294 GGCAACGGCA-ACCCCTTTGGGTCCCTCCCTTTGATGGAGGCTTTTCAATGTA 352

DB 241 GGCAACGGCA-ACCCCTTTGGGTCCCTCCCTTTGATGGAGGCTTTTCAATGTA 300

QY 353 TTTCACCTATTAATCTTGAACAGTCACTCTTGTGTCATAGTTTCTTAACGGCTTGAAC 412

DB 301 TTTCACCTATTAATCTTGAACAGTCACTCTTGTGTCATAGTTTCTTAACGGCTTGAAC 360

QY 413 TGAGCTTTGCTGCGCATCACCACCTGCTGTTTGGCGACACGACAGACCCGCGCTGACT 472

DB 361 TGAGCTTTGCTGCGCATCACCACCTGCTGTTTGGCGACACGACAGACCCGCGCTGACT 420

QY 473 CCCATCCCTTGATCATGACGAGGTGTC 501

DB 421 CCCATCCCTTGATCATGACGAGGTGTC 449

RESULT 8

ID ABN97946 standard; DNA; 711 BP.

XX ABN97946;

XX 01-AUG-2002 (first entry)

DE Human retroviral sequence R1.

KM Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

KM multiple sclerosis; ds.

OS Human retrovirus.

XX MO9967395-A1.

XX 29-DEC-1999.

PF 23-JUN-1999; 99MO-FR01513.

XX 23-JUN-1998; 98FR-0007920.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Alliel PM, Perin J, Rieger F;

DR WPI; 2000-160587/14.

XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
 PT used for diagnosis, treatment and prevention of autoimmune and
 PT neurological diseases -

XX Claim 3; Fig 3; 225pp; French.

XX The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention.

SQ Sequence 711 BP; 170 A; 204 C; 162 G; 175 T; 0 other;

Query Match 79.1%; Score 396.2; DB 21; Length 711;
 Best Local Similarity 95.3%; Pred. No. 2e-125;
 Matches 430; Conservative 0; Mismatches 18; Indels 3; Gaps 2;

QY 52 ACTGAGAGACAGACTAGCTGATTTCTTAGCTGACTAAGATCCCTAAGCTTACGCTG 111

DB 1 ACTGAGAGACAGACTAGCTGATTTCTTAGCTGACTAAGATCCCTAAGCTTACGCTG 60

QY 112 GAAGGTACCAACATCCACCTTAAACAGGGGCTTGCACTTAGCTCAGCTGACCAAT 171

DB 61 GAAGGTACCAACATCCACCTTAAACAGGGGCTTGCACTTAGCTCAGCTGACCAAT 120

QY 172 CAGAGAGCTCACTAAATGCTAATTAGCAAGAGAGTAAAGAAATAGCCATCATCT 231

DB 121 CAGAGAGCTCACTAAATGCTAATTAGCAAGAGAGTAAAGAAATAGCCATCATCT 180

QY 232 CTATTGCTGAGAGACAGAGGAGGACAAATGATCGGATATAAACCAGTCTTGAG 291

DB 181 CTATTGCTGAGAGACAGAGGAGGACAAATGATCGGATATAAACCAGTCTTGAG 240

QY 292 CCGGCAACGGCA-ACCCCTTTGGGTCCCTCCCTTTGATGGAGGCTTGTGTTTCAATGC 350


```

Db      547 TCACCTATTAATCTTCGACACTCTCTGCTGTGTGTTGTAAGGTTAGCTG 606
Qy      415 AGCTTTCCTGCGCCATCCACCACTCTGTTGCCJCCACCCGAGACCCGCGCTGATCC 474
Db      607 AGCTTTCCTGCGCCATCCACCACTCTGTTGCCJCCACCCGAGACCTGCGCTGATCC 666
Qy      475 CATCCCTCGATCATGCGAGGCTGTC 501
Db      667 CATCCCTCGATCATGCGAGGCTGTC 693

RESULT 10
AAS65964
ID      AAS65964 standard; cDNA; 893 BP.
XX
AC      AAS65964;
XX
DT      13-FEB-2002 (first entry)
XX
DE      DNA encoding novel human diagnostic protein #1768.
XX
KM      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM      food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS      Homo sapiens.
XX
PN      WO200175067-A2.
XX
PD      11-OCT-2001.
XX
PF      30-MAR-2001; 2001WO-US08631.
XX
PR      31-MAR-2000; 2000US-0540217.
PR      23-AUG-2000; 2000US-0649167.
XX
PA      (HYSE-) HYSEQ INC.
PI      Drmanac RT, Liu C, Tang YT;
XX
DR      WPI; 2001-639362/73.
DR      P-PSDB; ABE01777.
XX
PT      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity -
XX
PS      Claim 1; SEQ ID No 1768; 103pp; English.
XX
XX      The invention relates to isolated polynucleotide (I) and
XX      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      and gene mapping, and in recombinant production of (II). The
XX      polynucleotides are also used in diagnostics as expressed sequence tags
XX      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      to restore normal activity of (II) or to treat disease states involving
XX      (II). (II) is useful for generating antibodies against it, detecting or
XX      quantitating a polypeptide in tissue, as molecular weight markers and as
XX      a food supplement. (II) and its binding partners are useful in medical
XX      imaging of sites expressing (II). (I) and (II) are useful for treating
XX      disorders involving aberrant protein expression or biological activity.
XX      The polypeptide and polynucleotide sequences have applications in
XX      diagnostics, forensics, gene mapping, identification of mutations
XX      responsible for genetic disorders or other traits to assess biodiversity
XX      and to produce other types of data and products dependent on DNA and
XX      amino acid sequences. AAS64197-AAS94564 represent novel human
XX      diagnostic coding sequences of the invention.
XX      Note: The sequence data for this patent did not appear in the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ      Sequence 893 BP; 223 A; 264 C; 202 G; 204 T; 0 other;

```

```

Query Match      75.4%; Score 377.8; DB 23; Length 893;
Best Local Similarity 88.6%; Pred. No. 4,7e-119;
Matches 451; Conservative 0; Mismatches 37; Indels 21; Gaps 3;

Qy      4 TCGGCCAACCTCCCAACAGCACTTAGTTTCTCTGTGAGATGGGGGACGAGAGACAG 63
Db      188 TCGGCCAACCTCCCAACAGCACTTAGTTTCTCTGTGAGATGGGGGACGAGAGACAG 247
Qy      64 GACTAGCTGATTTCTTAGGCTGACTAAGATCCCTAAGCCTAGCTGGAGAGTACAC 123
Db      248 GACTAGCTGATTTCTTAGGCTGACTAAGATCCCTAAGCCTAGCTGGAGAGTACAC 307
Qy      124 ATCCACCTTTAACAACGAGGCTTSCAATTAGCTCACTGACCAATC-----AG 174
Db      308 ATCCACCTTTAACAACGAGGCTTSCAATTAGCTCACTGACCAATCAGTAAGAAAG 367
Qy      175 AGAGCTCACTAAATGCTAATTAAGCAAAAGACAGAGGTAAAGAAATGCAATCATCTA 234
Db      368 AGAGCCCGCTAAATGCTAATTAAGCAAAAGACAGAGGTAAAGAAATGCAATCATCTA 427
Qy      235 TTGCCTGAGACAGACAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCG 294
Db      428 TTGCCTGAGACAGACAGGAGGAGCAATGATCAGATATTAACCAAGCATTTGAGCCG 487
Qy      295 GCAAGGCAACCCCTTTGGGTCCCTCCCTTTGTATGAGAGCTGTGTTTCATGCTATT 354
Db      488 GCAAGCACTACCTCTTTGGGTCCCTCCCTTTGTATGAGAGAGCTGTG-----TT 537
Qy      355 TCACCTATTAATATCTTGCACTG--CACTCTTCTGCTCCATGTTTCTTAGGCTTAGC 412
Db      538 TCACCTATTAATATCTTGCACTGCACTCTTCTGCTGTGTGTTGTTAGGCTGAGC 597
Qy      413 TGAGCTTTCGCTCGCCATCCACCACTGCTGTTGCCGACCGACCGACCCGCTGACT 472
Db      598 TGAGCTTTCGCTCAACCGTCACCACTGCTGTTGCCACCGTTGCAAGACCAATCGTACT 657
Qy      473 CCCATCCCTCGATCATGACAGGCTGTC 501
Db      658 TCCACCCCTCCAGATATGACAGGCTGTC 686

RESULT 11
AAS55630
ID      AAS55630 standard; DNA; 2781 BP.
XX
AC      AAS55630;
XX
DT      29-MAY-2001 (first entry)
XX
DE      Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
KM      Envelope protein; HENV; syncytia formation; placental development;
KM      syncytia; cancer; cell adhesion; ss.
XX
OS      Human endogenous retrovirus.
XX
FH      Key      Location/Qualifiers
FT      CDS      762..2378
FT              /*tag= a
FT              /product= "envelope protein"
XX
PN      WO200116171-A1.
XX
PD      08-MAR-2001.
XX
PF      01-SEP-2000; 2000WO-FR02429.
XX
PR      01-SEP-1999; 99FR-0011141.
PR      15-SEP-1999; 99FR-0011793.
XX
PA      (INNR ) BIO MERIEUX.
PA      (INNR ) INST NAT SANTE & RECH MEDICALE.

```


XX Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
PI WPI: 2001-226676/23.
XX P-PSDB; AAB67652.
DR
XX
PT Detecting expression of human endogenous retrovirus envelope protein in
PT cells of a tissue or culture, from its ability to induce syncytia -
XX
XX Disclosure; Page 44-45; 57pp; French.
XX
CC The present sequence encodes a human endogenous retrovirus envelope
CC protein. The specification describes a method for detecting expression
CC of an envelope protein from a human endogenous retrovirus (HERV), in
CC cells, of a tissue or culture. The method comprises detecting syncytia
CC formation due to the fusogenic properties of the envelope protein.
CC Envelope polypeptides and polynucleotides are used to produce
CC therapeutic or prophylactic compositions, particularly for treatment of
CC cancer, to correct defects in placental development (or other natural
CC formation of other types of syncytia), and to promote adhesion of cells
CC in grafts or cellular repair processes. Expression of sequences
CC antisense to the polynucleotide are used to prevent formation of
CC syncytia.
XX
SQ Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 other;
Query Match 75.2%; Score 377; DB 22; Length 2781;
Best Local Similarity 100.0%; Pred. No. 1,6e-118; Indels 0; Gaps 0;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 TCGGCAACCTCCCAAGACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 63
DB 2387 TCGGCAACCTCCCAAGACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 2446
QY 64 GACTAGCTGATTTCTTCTAGGCTGATAGAAATCCCTAAGCTGAGGAGGTACAC 123
DB 2447 GACTAGCTGATTTCTTCTAGGCTGATAGAAATCCCTAAGCTGAGGAGGTACAC 2506
QY 124 ATCCACCTTTAAACAGGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 183
DB 2507 ATCCACCTTTAAACAGGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 2566
QY 184 TAAATGCTAATTAGGCAAGAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTTGG 243
DB 2567 TAAATGCTAATTAGGCAAGAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTTGG 2626
QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGGAGCCGCAACGGCA 303
DB 2627 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGGAGCCGCAACGGCA 2686
QY 304 ACCGCTTTGGGTCGCCCTCTTGTATGGGAGCTCTGTTTCATGCTATTCTACTCTAT 363
DB 2687 ACCGCTTTGGGTCGCCCTCTTGTATGGGAGCTCTGTTTCATGCTATTCTACTCTAT 2746
QY 364 TAAATCTTGAACCTGCA 380
DB 2747 TAAATCTTGAACCTGCA 2763
RESULT 12
AA77526
ID AA77526 standard; cDNA; 2946 BP.
XX
AC AA77526;
XX
DT 10-AUG-1999 (first entry)
DE Human secreted protein AJ172_2 cDNA.
XX
KW Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;

KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
XX Homo sapiens.
OS
XX W09926972-A1.
XX
XX 03-JUN-1999.
XX
XX 17-NOV-1998; 98WO-US24614.
XX
XX 20-OCT-1998; 98US-0175928.
XX 21-NOV-1997; 97US-0976110.
XX 18-MAY-1998; 98US-0080478.
XX
XX (GENY) GENETICS INST INC.
XX
XX Collins-Racie LA, Evans C, Jacobs K, Lavallie ER;
PI McCoy JM, Werberg D, Treacy M;
XX
XX WPI: 1999-357813/30.
DR P-PSDB; AAY08622.
XX
XX New polynucleotides encoding secreted proteins
XX
XX Claim 13a; Page 100-101; 142pp; English.
XX
CC This invention describes novel human secreted proteins encoded by
CC polynucleotides isolated from human adult testes, adult brain, adult
CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC libraries. The products of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. The polynucleotides are also stated to be useful
CC for gene therapy.
XX
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;
Query Match 74.9%; Score 375.4; DB 20; Length 2946;
Best Local Similarity 99.7%; Pred. No. 6e-118; Indels 1; Gaps 0;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 TCGGCAACCTCCCAAGACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 63
DB 2553 TCGGCAACCTCCCAAGACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 2612
QY 64 GACTAGCTGATTTCTTCTAGGCTGATAGAAATCCCTAAGCTGAGGAGGTACAC 123
DB 2613 GACTAGCTGATTTCTTCTAGGCTGATAGAAATCCCTAAGCTGAGGAGGTACAC 2672
QY 124 ATCCACCTTTAAACAGGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 183
DB 2673 ATCCACCTTTAAACAGGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 2732
QY 184 TAAATGCTAATTAGGCAAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTTGG 243
DB 2733 TAAATGCTAATTAGGCAAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTTGG 2792
QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGGAGCCGCAACGGCA 303
DB 2793 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGGAGCCGCAACGGCA 2852
QY 304 ACCGCTTTGGGTCGCCCTCTTGTATGGGAGCTCTGTTTCATGCTATTCTACTCTAT 363
DB 2853 ACCGCTTTGGGTCGCCCTCTTGTATGGGAGCTCTGTTTCATGCTATTCTACTCTAT 2912

Qy 364 TAAATCTTGCAACTGCA 380
Db 2913 TAAATCTTGCAACTGCA 2929

RESULT 13

AA259468
AA259468 standard; cDNA; 2946 BP.

AA259468;

11-APR-2000 (first entry)

Human secreted protein AJ172_2 polynucleotide sequence.

Human; secreted protein; disease diagnosis; pre-eclampsia; cancer; placental pathology; metastasis inhibition; nutritional activity; immune stimulator; haematopoiesis regulator; tissue growth; tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115; gene therapy; ss.

Homo sapiens.

WO9960020-A1.

25-NOV-1999.

17-MAY-1999; 99WO-US10915.

18-MAY-1998; 98US-0080478.

20-OCT-1998; 98US-0175928.

(GENY) GENETICS INST INC.

Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

Metberg D, Mi S, Treacy M;

WPI; 2000-116311/10.

P-PSDB; AAY67313.

New polynucleotides encoding secreted cDNA libraries, used to develop products for the diagnosis and treatment of neoplastic disease

Claim 14; Page 107-108; 149pp; English.

This is the human secreted protein AJ172_2 nucleotide sequence, obtained from a human adult testes cDNA library. The invention relates to secreted human and murine proteins. The polynucleotides and proteins are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Detection of the levels of the proteins can be used for the diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate the expression or function of the proteins may be used for treating a neoplastic disease and inhibiting metastasis. Other suggested activities include nutritional activity (e.g. in feeds), cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotide sequences are also stated to be useful for gene therapy.

Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;

Query Match 74.9%; Score 375.4; DB 21; Length 2946;

Best Local Similarity 99.7%; Pred. No. 66-118;

Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

4 TCGGCCAACCTCCCAACGACCTAGCTTTCCTGATGAGGCGACTGAGACAG 63
2553 TCGGCCAACCTCCCAACGACCTAGCTTTCCTGATGAGGCGACTGAGACAG 2612

Qy 64 GACTAGCTGATTTCTGAGGCTGACTAAGATCCCTAAGCTTACTGGAGAGTGACCAC 123
Db 2613 GACTAGCTGATTTCTGAGGCTGACTAAGATCCCTAAGCTTACTGGAGAGTGACCAC 2672
Qy 124 ATCCACCTTTAAACAGGGGGCTTCACTTATGCTCACACCTGACCAATCAGAGAGCTCAC 183
Db 2673 ATCCACCTTTAAACAGGGGGCTTCACTTATGCTCACACCTGACCAATCAGAGAGCTCAC 2732
Qy 184 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCAAATCATATTGCTCTAG 243
Db 2733 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCAAATCATATTGCTCTAG 2792
Qy 244 AGCAGCAGAGAGGACATATATGCGATATAAACCAAGTCTTGAGCCGCGACAGCGCA 303
Db 2793 AGCAGCAGAGAGGACATATATGCGATATAAACCAAGTCTTGAGCCGCGACAGCGCA 2852
Qy 304 ACCCCCTTGGGTCCTCCCTTGTATGGAGCTCTGTTTCAATGCTATTCTACTAT 363
Db 2853 ACCCCCTTGGGTCCTCCCTTGTATGGAGCTCTGTTTCAATGCTATTCTACTAT 2912
Qy 364 TAAATCTTGCAACTGCA 380
Db 2913 TAAATCTTGCAACTGCA 2929

RESULT 14

AA24195
AAD24195 standard; cDNA; 2930 BP.

AA24195;

07-MAY-2002 (first entry)

Human syncytin cDNA.

Human; syncytin; pre-eclampsia; gestational trophoblast disorder;

choriocarcinoma; hydatiform mole; placental site tumour; abortion;

envelope gene; human endogenous defective retrovirus; HBRV-W; ss.

Homo sapiens.

Key Location/Qualifiers

FT CDS /tag= a

FT /product= "Syncytin"

MO200204678-A2.

17-JAN-2002.

09-JUL-2001; 2001WO-US21719.

07-JUL-2000; 2000US-216657P.

(GENY) GENETICS INST INC.

Keith JC, McCoy JM, Mi S;

WPI; 2002-171727/22.

P-PSDB; AAE14540.

Identifying a compound for treating a subject with or at risk of developing pre-eclampsia, comprises determining whether the expression or activity of syncytin in the cell is modulated in the presence of a test compound

Disclosure; Page 39-42; 43pp; English.

The invention relates to identifying compounds which are modulators of syncytin expression. The syncytin modulators are useful in diagnosis and treatment of pre-eclampsia and gestational trophoblast disorders (e.g. choriocarcinoma, hydatiform mole, placental site tumour and missed/incomplete abortion). Syncytin is a human gene derived from the

CC envelope gene of human endogenous defective retrovirus, HERV-W. The
CC present invention is based partly on the discovery that syncytin
CC expression is dramatically reduced in preclampsia, and is also
CC mis-localised to the apical syncytiotrophoblast membrane. The present
CC sequence is human syncytin cDNA.

XX Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 other;

Query Match 74.7%; Score 374.4; DB 24; Length 2930;
Best Local Similarity 99.7%; Pred. No. 1.3e-117;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCGGCAACTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGACTGAGAGACAG 63
Db TCGGCCAAGCTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGACTGAGAGACAG 2614
QY 64 GACTAGCTGATTTCTTGAGCTGACTAAGATCCCTAAGCCCTAGCTGGGAAGTACAC 123
Db 2615 GACTAGCTGATTTCTTGAGCTGACTAAGATCCCTAAGCCCTAGCTGGGAAGTACAC 2674
QY 124 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 183
Db 2675 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 2734
QY 184 TAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGAG 243
Db 2735 TAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGAG 2794
QY 244 AGCAGACAGAGAGGACATGATCGGATATTAACCAAGTCTTGAGACCGGCAACGGCA 303
Db 2795 AGCAGACAGAGAGGACATGATCGGATATTAACCAAGTCTTGAGACCGGCAACGGCA 2854
QY 304 ACCCCCTTGGGTCCCTCCCTTTGATGAGAGCTGTGTTTCAATGCTATTACCTCAT 363
Db 2855 ACCCCCTTGGGTCCCTCCCTTTGATGAGAGCTGTGTTTCAATGCTATTACCTCAT 2914
QY 364 TAAATCTTGCAACTGC 379
Db 2915 TAAATCTTGCAACTGC 2930

RESULT 15
AAK25665
ID AAK25665 standard; cDNA to mRNA; 7582 BP.

XX AAK25665;

XX 21-MAY-1999 (first entry)

XX Complete human endogenous retrovirus W genome.

XX Clone; human endogenous retrovirus; genome; autoimmune disease;
XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
XX disseminated lupus erythematosus; pregnancy; chromosomal marker; 89.

XX Human endogenous retrovirus.

XX MO9902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-FR01442.

XX 07-JUL-1997; 97FR-0008815.

XX (INMR) BIO MERIEUX.

XX Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;

XX WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -
XX expressed exclusively in placenta and useful in diagnosis and
XX

PT therapy of autoimmune disease, and abnormal or failed pregnancy
XX
XX Claim 1; Page 71-74; 106pp; French.

CC This sequence represents the complete sequence of the human endogenous
CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC peptides encoded by them are markers of autoimmune disease (e.g. multiple
CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC insulin-dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility.

XX Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;

Query Match 74.5%; Score 373.4; DB 20; Length 7582;
Best Local Similarity 97.6%; Pred. No. 4.9e-117;
Matches 368; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCAACTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGACTGAGAGACAG 63
Db 7206 TCGGCCAAGCTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGACTGAGAGACAG 7265
QY 64 GACTAGCTGATTTCTTGAGCTGACTAAGATCCCTAAGCCCTAGCTGGGAAGTACAC 123
Db 7266 GACTAGCTGATTTCTTGAGCTGACTAAGATCCCTAAGCCCTAGCTGGGAAGTACAC 7325
QY 124 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 183
Db 7326 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 7385
QY 184 TAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGAG 243
Db 7386 TAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGAG 7445
QY 244 AGCAGACAGAGAGGACATGATCGGATATTAACCAAGTCTTGAGACCGGCAACGGCA 303
Db 7446 AGCAGACAGAGAGGACATGATCGGATATTAACCAAGTCTTGAGACCGGCAACGGCA 7505
QY 304 ACCCCCTTGGGTCCCTCCCTTTGATGAGAGCTGTGTTTCAATGCTATTACCTCAT 363
Db 7506 ACCCCCTTGGGTCCCTCCCTTTGATGAGAGCTGTGTTTCAATGCTATTACCTCAT 7565
QY 364 TAAATCTTGCAACTGCA 380
Db 7566 TAAATCTTGCAACTGCR 7582

Search completed: April 17, 2003, 05:07:30
Job time : 206.787 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:35:35 ; Search time 34.6897 Seconds
(without alignments)
4429.120 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000
Perfect score: 501
Sequence: 1 gtctcgccacaccccccacccca.....ctggatcatgcaggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	375.4	74.9	2946	US-09-175-928-3	Sequence 3, Appli
2	238.4	47.6	279	US-08-686-878A-50	Sequence 50, Appli
3	238.4	47.6	279	US-08-721-489-4	Sequence 4, Appli
4	46.8	9.3	7218	US-08-233-463-14	Sequence 14, Appli
5	34.2	6.8	80246	US-09-078-294-4	Sequence 3, Appli
6	34.2	6.8	80595	US-09-078-294-3	Sequence 3, Appli
7	34	6.8	7218	US-08-233-463-14	Sequence 14, Appli
8	33	6.6	1859	US-08-691-563C-46	Sequence 46, Appli
9	32	6.4	1417	US-08-117-362-31	Sequence 31, Appli
10	32	6.4	1417	US-08-486-924-31	Sequence 31, Appli
11	32	6.4	1722	US-08-691-563C-58	Sequence 58, Appli
12	30.8	6.1	9763	US-08-973-273-1	Sequence 1, Appli
13	30.6	6.1	11827	US-09-733-455-3	Sequence 3, Appli
14	30.2	6.0	1043	US-09-422-576D-6	Sequence 6, Appli
15	30.2	6.0	1091	US-09-422-576D-5	Sequence 5, Appli
16	30.2	6.0	1165	US-09-422-576D-5	Sequence 25, Appli
17	30.2	6.0	1831	US-09-422-576D-1	Sequence 1, Appli
18	29.4	5.9	29598	US-09-341-587-6	Sequence 6, Appli
19	29.2	5.8	1289	US-08-896-164-16	Sequence 16, Appli
20	29.2	5.8	1132	US-08-894-731-3	Sequence 3, Appli
21	29	5.8	2023	US-08-961-083-199	Sequence 199, App
22	28.8	5.7	1066	US-08-793-559-2	Sequence 2, Appli
23	28.8	5.7	1066	US-08-793-559-2	Sequence 2, Appli
24	28.8	5.7	1066	US-08-793-559-2	Sequence 3, Appli
25	28.8	5.7	1066	US-08-793-559-2	Sequence 3, Appli
26	28.8	5.7	1066	US-08-793-559-2	Sequence 1, Appli
27	28.8	5.7	1066	US-08-793-559-2	Sequence 2, Appli

28	28.6	5.7	3552	US-09-157-210-3	Sequence 3, Appli
29	28.4	5.7	6836	US-08-976-259-73	Sequence 73, Appli
30	28.2	5.6	623	US-09-385-982-52	Sequence 52, Appli
31	28.2	5.6	16389	US-09-741-154-3	Sequence 3, Appli
32	28.2	5.6	19307	US-08-836-022A-10	Sequence 10, Appli
33	28.2	5.6	19307	US-09-427-048A-10	Sequence 10, Appli
34	28	5.6	2364	US-08-838-219B-5	Sequence 5, Appli
35	28	5.6	2364	US-08-838-219B-5	Sequence 5, Appli
36	28	5.6	2364	US-09-233-752A-5	Sequence 5, Appli
37	28	5.6	2364	US-09-402-036-5	Sequence 5, Appli
38	28	5.6	2364	US-09-904-226-5	Sequence 5, Appli
39	28	5.6	2370	US-09-002-285-79	Sequence 79, Appli
40	28	5.6	2370	US-09-002-285-95	Sequence 95, Appli
41	28	5.6	2374	US-09-002-285-97	Sequence 97, Appli
42	28	5.6	2375	US-08-960-780-3	Sequence 3, Appli
43	28	5.6	2375	US-08-960-780-7	Sequence 7, Appli
44	28	5.6	2375	US-09-073-898-3	Sequence 3, Appli
45	28	5.6	2375	US-09-073-898-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-175-928-3
; Sequence 3, Application US/09175928A
; Patent No. 6312921

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M.

APPLICANT: Lavalite, Edward R.

APPLICANT: Collins-Racie, Lisa A.

APPLICANT: Evans, Cheryl

APPLICANT: Merberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Mi, Sha

APPLICANT: Genetics Institute, Inc.

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

FILE REFERENCE: 6006B A1172A

CURRENT APPLICATION NUMBER: US/09175, 928A

CURRENT FILING DATE: 1998-10-20

NUMBER OF SEQ ID NOS: 62

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3

LENGTH: 2946

TYPE: DNA

ORGANISM: Homo sapiens

US-09-175-928-3

Query Match Best Local Similarity 74.9%; Score 375.4; DB 4; Length 2946;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	4	TCGGCAACTCCCAACAGCACTAGGTTTCCTGTTGAAGTGGGAGACTAGAGACG	63
DB	2553	TCGGCAACTCCCAACAGCACTAGGTTTCCTGTTGAAGTGGGAGACTAGAGACG	2612
QY	64	GACTGCTGATTTCTTGCTGACTTAAGATCCTTAAGCTGAGTGGAGGTGACAC	123
DB	2613	GACTGCTGATTTCTTGCTGACTTAAGATCCTTAAGCTGAGTGGAGGTGACAC	2672
QY	124	ATCCACCTTTAAACGCGGGCTTGCACTTGTGCACCTGACCAATCAGAGACTCAC	183
DB	2673	ATCCACCTTTAAACGCGGGCTTGCACTTGTGCACCTGACCAATCAGAGACTCAC	2732
QY	184	TAAATGCTTAATTAGCAAGAAGAGAGATTAAGAAATAGCAATCATCTATTGCTGAG	243
DB	2733	TAAATGCTTAATTAGCAAGAAGAGAGATTAAGAAATAGCAATCATCTATTGCTGAG	2792
QY	244	AGCAGAGAGAGAGAGAGATGATGCGGATATTAACCCAAAGTCTTGAACCGGCAACGCGCA	303
DB	2793	AGCAGAGAGAGAGAGAGATGATGCGGATATTAACCCAAAGTCTTGAACCGGCAACGCGCA	2852

Qy 304 ACCCCCTTTGGGTCCTCCCTTTGTGATGGAGCTCTGTGTTTCATGCTATTTCACCTAT 363
Db 2853 ACCCCCTTTGGGTCCTCCCTTTGTGATGGAGCTCTGTGTTTCATGCTATTTCACCTAT 2912
Qy 364 TAAATCTTGCACTGCA 380
Db 2913 TAAATCTTGCACTGCA 2929

RESULT 2

US-08-686-878A-50
; Sequence 50, Application US/08686878A
; Patent No. 5708157
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallee, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Werberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/686,878A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-686-878A-50

Query Match 47.6%; Score 238.4; DB 1; Length 279;
Best Local Similarity 92.4%; Pred. No. 8.2e-72;
Matches 242; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

Qy 119 ACCACATCCACCTTTAAACACGGGGCTTGCACTTACCTCACTGACCTGACCAATCAGAGAG 178
Db 1 RCACATCCACCTTTAAACACGGGGCTTGCAANAAAGATNACCTTGACCAATCAGAGAG 60
Qy 179 CTCACCTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCTATTTCG 238
Db 61 NTCANTAAATATATATTGCGCAAAAACAGAGGTAAAGAAATAGCCATCTATTTCG 120
Qy 239 CTGAGAGCAGACAGAGAGGACATATCGGATATATAACCAAGCTTCGAGCCGGCAA 298
Db 121 CTGAGAGCAGACAGAGAGGACATATCGGATATATAACCAAGCTTTNGAGCCGGCAA 180
Qy 299 CGGCAACCCCTTTGGGTCCCTCCTTTGTATGGAGCTCTGTGTTTCATGCTATTTCAC 358

Db 181 |||||
Qy 359 TCTATTAATCTTGCACTGCA 380
Db 241 TTTATTAATTTTGCACTGCA 262

RESULT 3

US-08-721-489-4
; Sequence 4, Application US/08721489
; Patent No. 5786465
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallee, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Werberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,489
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-721-489-4

Query Match 47.6%; Score 238.4; DB 1; Length 279;
Best Local Similarity 92.4%; Pred. No. 8.2e-72;
Matches 242; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

Qy 119 ACCACATCCACCTTTAAACACGGGGCTTGCACTTACCTCACTGACCTGACCAATCAGAGAG 178
Db 1 RCACATCCACCTTTAAACACGGGGCTTGCAANAAAGATNACCTTGACCAATCAGAGAG 60
Qy 179 CTCACCTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCTATTTCG 238
Db 61 NTCANTAAATATATATTGCGCAAAAACAGAGGTAAAGAAATAGCCATCTATTTCG 120
Qy 239 CTGAGAGCAGACAGAGAGGACATATCGGATATATAACCAAGCTTCGAGCCGGCAA 298
Db 121 CTGAGAGCAGACAGAGAGGACATATCGGATATATAACCAAGCTTTNGAGCCGGCAA 180
Qy 299 CGGCAACCCCTTTGGGTCCCTCCTTTGTATGGAGCTCTGTGTTTCATGCTATTTCAC 358
Db 181 CGGCAACCCCTTTGGGTCCCTCCTTTGTATGGAGCTCTGTGTTTCATGCTATTTCAN 240

RESULT 7
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ19pt-F18
US-08-232-463-14

Query Match 6.8%; Score 34; DB 1; Length 7218;
Best Local Similarity 7.1%; Pred. No. 0.3;
Matches 22; Conservative 155; Mismatches 135; Indels 0; Gaps 0;

Qy 42 GAGATGGGGAGCTGAGACAGAGCTAGTTCTGAGCTGACTAAGATCCCTAA 101
Db 1332 RRR 1273
Qy 102 GCTAGCTGGGAGGAGCAACATCCACTTAAACGCGGGTTCACCTAGCTACA 161
Db 1272 RRR 1213
Qy 162 CCTGACCAATCAGAGAGCTAATAATGCTAATATGCAAGACAGAGTAAAGAAAT 221
Db 1212 RRR 1153
Qy 222 AGCCATCATCTATTGCTGAGACACAGAGGAGGACAGATGCGGATATAACCA 281
Db 1152 RRR 1093
Qy 282 AGCTTCGAGCGGAGCAAGGCAACCCCTTTGGTCCCTCCCTTTGTATGGAAGCTCG 341
Db 1092 RRR 1033

Qy 342 TTTTCATGCTAT 353
Db 1032 AATTAATCTGT 1021

RESULT 8
US-08-691-563C-46
Sequence 46, Application US/08691563C
Patent No. 6001987
GENERAL INFORMATION:
APPLICANT: HERVE PERRON
APPLICANT: Frederic BESME
APPLICANT: Frederic BEDIN
APPLICANT: Glaucia PARANHOS-BACCALA
APPLICANT: Florence KOMURIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1859 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-691-563C-46

Query Match 6.6%; Score 33; DB 3; Length 1859;
Best Local Similarity 6.2%; Pred. No. 0.34; 20; Indels 0; Gaps 0;
Matches 45; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 14 TCCCAACAGACACTTAGTTTCTGTTGAGATGGGGAGCTGAGACAGACTAGCTGG 73
Db 1687 TTCCCAAGACAGCTGGGGGTGTCCTTTAGAGTGGGATGAGAGGAGCCAGCTGG 1746
Qy 74 ATTTC 78
Db 1747 ACTTC 1751

RESULT 9
US-08-117-362-31/c
Sequence 31, Application US/08117362
Patent No. 5595872
GENERAL INFORMATION:
APPLICANT: Metterau II, John R.
APPLICANT: Sharp, Daru Y.
APPLICANT: Gregg, Richard E.


```

; TITLE OF INVENTION: MICROSMAL TRIGLYCERIDE TRANSFER PROTEIN
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/117,362
; FILING DATE: 03-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; REFERENCE/DOCKET NUMBER: DC21b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-5901
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 783..890
; US-08-117-362-31

Query Match      6.4%; Score 32; DB 1; Length 1417;
Best Local Similarity 45.6%; Pred. No. 0.6;
Matches 154; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 71 TGAATTCCTAGAGCTGACTAAGAAATCCCTAAGCCTAGCTGGAAAGTGACCAATCCACC 130
Db 1345 TGCCTTTTTCATTGGGCTAATCTCTTGAAGTCAATGCTAAGCAATAAAGAAA 1286
QY 131 TTTAAACACGGGGCTTGCACTTAGCTACACCTGACCAATCAGAGCTCACTAAATG 190
Db 1285 TTTAAGAACTGGGTTTTTAAAAAAATCCTCATTTACAAATGAAGAGCTGCATT 1226
QY 191 CTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCAATC---ATCTATTGCTGAGAGC 246
Db 1225 CAATGTACCAACAACAGCAGTAAATAGTGGTATCTTTTTCCTTTTAAAAAC 1166
QY 247 ACAGCAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTCGAGCCGGCAAGCAACC 306
Db 1165 AGACCAATGAATGAAGAAATGAAGCAAGAACACATGCTGAGGAACCTGAATAAAGA 1106
QY 307 CCCTTGGGTCCTCCCTCTTGTATGGAGCTGTGTTTCAATGCTATTTCACTATTTAA 366
Db 1105 AATATCTGTTTTCTTAATGTTTGAAGGATTTGAATTTGGTTTCATTTCACCTACGCA 1046
QY 367 ATCTGCACTGCACTCTTGTGTCATGTTCTTACG 404
Db 1045 TGCCTTCACAAAACATTCCAAGTTACATGTTTGAAG 1008

RESULT 10
US-08-486-924-31/c
; Sequence 31, Application US/08486924
; Patent No. 5789197
; GENERAL INFORMATION:
; APPLICANT: Wetters II, John R.
; APPLICANT: Sharp, Daru Y.

```

```

; APPLICANT: Gregg, Richard E.
; TITLE OF INVENTION: MICROSMAL TRIGLYCERIDE TRANSFER PROTEIN
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burton Rodney
; STREET: P.O. Box 4000
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,924
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/117,362
; FILING DATE: 03-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaul, Timothy J.
; REGISTRATION NUMBER: 33,111
; REFERENCE/DOCKET NUMBER: DC21b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-5901
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 783..890
; US-08-486-924-31

Query Match      6.4%; Score 32; DB 1; Length 1417;
Best Local Similarity 45.6%; Pred. No. 0.6;
Matches 154; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 71 TGAATTCCTAGAGCTGACTAAGAAATCCCTAAGCCTAGCTGGAAAGTGACCAATCCACC 130
Db 1345 TGCCTTTTTCATTGGGCTAATCTCTTGAAGTCAATGCTAAGCAATAAAGAAA 1286
QY 131 TTTAAACACGGGGCTTGCACTTAGCTACACCTGACCAATCAGAGCTCACTAAATG 190
Db 1285 TTTAAGAACTGGGTTTTTAAAAAAATCCTCATTTACAAATGAAGAGCTGCATT 1226
QY 191 CTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCAATC---ATCTATTGCTGAGAGC 246
Db 1225 CAATGTACCAACAACAGCAGTAAATAGTGGTATCTTTTTCCTTTTAAAAAC 1166
QY 247 ACAGCAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTCGAGCCGGCAAGCAACC 306
Db 1165 AGACCAATGAATGAAGAAATGAAGCAAGAACACATGCTGAGGAACCTGAATAAAGA 1106
QY 307 CCCTTGGGTCCTCCCTCTTGTATGGAGCTGTGTTTCAATGCTATTTCACTATTTAA 366
Db 1105 AATATCTGTTTTCTTAATGTTTGAAGGATTTGAATTTGGTTTCATTTCACCTACGCA 1046
QY 367 ATCTGCACTGCACTCTTGTGTCATGTTCTTACG 404
Db 1045 TGCCTTCACAAAACATTCCAAGTTACATGTTTGAAG 1008

RESULT 11
US-08-691-563C-58
; Sequence 58, Application US/08691563C

```

Patent No. 6001987
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Frederic BESEME
APPLICANT: Frederic BBDIN
APPLICANT: Glaucia PARANHOS-BACCALA
APPLICANT: Florence KOMURIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1722 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-691-563C-58

Query Match 6.4%; Score 32; 1B 3; Length 1722;
Best Local Similarity 73.2%; Pred. No. 0.66;
Matches 41; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 GTCCTGCGCCAACTCCCAACAGCACTTAGTTCCTGTTGAGATGGGGACTGA 56
Db 1024 GTTCATCGCCAAATTCACCAACGACGATGGGGTGTCTGTTTAGAGGGGGGATGGA 1079

RESULT 12
US-08-973-273-1/C
Sequence 1, Application US/08973273
Patent No. 6140085
GENERAL INFORMATION:
APPLICANT: Dean, Caroline
APPLICANT: MacKnight, Richard C
APPLICANT: Bancroft, Ian
APPLICANT: Lister, Clare K
TITLE OF INVENTION: Genetic Control of Flowering
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye P.C.
STREET: 1100 No. 6140085th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,273
FILING DATE: 01-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB96/01332
FILING DATE: 03-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 951196.9
FILING DATE: 02-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-29
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9763 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Arabidopsis thaliana
STRAIN: Columbia
US-08-973-273-1

Query Match 6.1%; Score 30.8; DB 3; Length 9763;
Best Local Similarity 63.5%; Pred. No. 4.4;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 162 CCTGACCAATGAGAGCTCACTTAATGCTTAATTAGGCAAGCAGAGGTAAGAAT 221
Db 4845 CCTCAGAAAGAAAGAACTATATACCTGATGTAAGCAAAAGAAACAGGAAAGAAA 4786

Qy 222 AGCCAAATCATCTAT 235
Db 4785 GGATCAACATCCAT 4772

RESULT 13
US-09-739-455-3
Sequence 3, Application US/09739455
Patent No. 6413756
GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CLO00653
CURRENT APPLICATION NUMBER: US/09/739,455
CURRENT FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 11827
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(11827)
OTHER INFORMATION: n = A,T,C or G
US-09-739-455-3

Query Match 6.1%; Score 30.6; DB 4; Length 11827;
Best Local Similarity 51.1%; Pred. No. 5.8;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY	103	CTTAGCTGGGAAGTGAGCACAACATCCACCCTTTAAACACGGGCGCTTGCAACTTAGCTACACG	162
Db	10853	CCCACTGCTGGAGTCGGAACTGCTGCCCTTTGTTGGCGGCGCTTGTTCTTAAATCAGTT	10912
QY	163	CTGACCCATTCAGAGGCTCATCAAAATAGTATATTGGCGAAACACAGAGGTAAAGAAATTA	222
Db	10913	CCCTCTTAGAGATTATTCACATCAAAAAAAATTTAGTTTTGAAAAAGAAATTGAGAAATTA	10972
QY	223	GCCATTCATCTATTGGCCCTGAG	243
Db	10973	CAGAAACATGAATTTCACGAG	10993

```

RESULT 14
US-09-422-576D-6
; Sequence 6, Application US/09422576D
; Patent No. 6395549
; GENERAL INFORMATION:
; APPLICANT: Tuan, Dorothy
; APPLICANT: Long, Qiaoming
; APPLICANT: Bengra, Chikn
; TITLE OF INVENTION: long Terminal Repeat, Enhancer, and Insulator Sequences for Use i
; TITLE OF INVENTION: Recombinant Vectors
; FILE REFERENCE: M0351-205010
; CURRENT APPLICATION NUMBER: US/09/422,576D
; CURRENT FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/105,256
; PRIOR FILING DATE: 1998-10-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-422-576D-6

```

	Query Match	Score	DB	Length
Best Local Similarity	62.1%	6.0%	30.2	1043
Matches	64	Conservative	0	Mismatches 38; Indels 1; Gaps 1

```

RESULT 15
US-09-422-576D-5
: Sequence 5, Application US/09422576D
: Patent No. 6395549
:
: GENERAL INFORMATION:
: APPLICANT: Tuan, Dorothy
: APPLICANT: Long, Xiaoming
: APPLICANT: Bengira, Chikh
: TITLE OF INVENTION: Long Terminal Repeat, Enhancer, and Insulator Sequences for Use in
: TITLE OF INVENTION: Recombinant Vectors
: FILE REFERENCE: M0351-205010
: CURRENT FILING DATE: 1999-10-21
: CURRENT APPLICATION NUMBER: US/09/422,576D
: PRIOR APPLICATION NUMBER: US 60/105,256
: PRIOR FILING DATE: 1998-10-22
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 5
: LENGTH: 1091
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus Sequence

```

```

; NAME/KEY: misc feature
; LOCATION: (437)..(437)
; OTHER INFORMATION: n = any nucleotide
US-09-422-576D-5

```

	Query Match	Best Local Similarity	Score	DB	Length
	6.0%	62.1%	30.2	4	1091
			Pred. No. 2.1		
	Matches 64	Conservative 0	Mismatches 38	Indels 1	Gaps 1
QY	253	GGAGGACATGATGGGATATAAACCAAGCTTCGAGCCGGCAACGCCCTTT	312		
DB	534	GGTGGGCCAGTTAAGAGATATAAAGCAGCGCTGCCCGAGCCAGCAGTGGCAACCGC-TC	592		
QY	313	GGGTCCTCTCCTTTGATGTGGAGCTCTGTTTCAATGCTATTT	355		
DB	553	GGGTCCCTTCACACTGTGGAAGCTTTGTTCTTTCGCTCTTT	635		

Search completed: April 17, 2003, 07:39:12
Job time : 104.69 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:57:20 ; Search time 123.415 Seconds
(without alignments)
4077.979 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 1 gtctcgagcaaccctccca.....ctgatcatgaggtgtcc 501

Sequence: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 639749 seqs, 50280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCIT_NEM_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCITUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	426	85.0	1329	7 US-08-979-847-108	Sequence 108, App
2	375.4	74.9	2946	9 US-10-114-893-134	Sequence 134, App
3	374.4	74.7	2930	10 US-09-902-535-1	Sequence 1, Appli
4	341.8	68.2	635	7 US-08-979-847-102	Sequence 102, App
5	317.6	63.4	180557	12 US-10-003-806-6	Sequence 6, Appli
6	317.6	63.4	180557	12 US-10-003-806-6	Sequence 9, Appli
7	311.4	62.2	410	10 US-09-880-107-538	Sequence 538, App
8	305.2	60.9	326014	10 US-09-731-231A-3	Sequence 3, Appli
9	286.6	57.2	541	10 US-09-864-761-8173	Sequence 8173, Ap
10	281.2	56.1	1894	10 US-09-864-761-4444	Sequence 4444, Ap
11	266.8	53.3	15425	9 US-10-091-504-1654	Sequence 1654, Ap
12	266.8	53.3	15425	9 US-10-091-504-1654	Sequence 1654, Ap
13	238.4	47.6	279	12 US-10-040-916-50	Sequence 50, Appli
14	203.8	40.7	569	10 US-09-864-761-14951	Sequence 14951, A
15	146.2	29.2	246	10 US-09-864-761-20462	Sequence 20462, A
16	136.8	27.3	446	10 US-09-811-284-3	Sequence 3, Appli
17	136.2	27.2	504	10 US-09-864-761-7027	Sequence 7027, Ap
18	129.6	25.9	3524	10 US-09-972-724-1	Sequence 1, Appli
19	124.8	24.9	475	10 US-09-864-761-895	Sequence 895, App

C 20	122.4	24.4	559	10 US-09-864-761-7501	Sequence 7501, Ap
C 21	111.6	22.3	440	10 US-09-864-761-3694	Sequence 3694, Ap
C 22	105.4	21.0	409	10 US-09-864-761-4153	Sequence 4153, Ap
C 23	92.2	18.4	579	10 US-09-864-761-13678	Sequence 13678, A
C 24	58.8	11.7	387	10 US-09-864-761-30194	Sequence 30194, A
C 25	58	11.6	525	10 US-09-893-737-31	Sequence 31, Appli
C 26	55.4	11.1	88	10 US-09-864-761-20907	Sequence 20907, A
27	50.6	10.1	275	10 US-09-864-761-17675	Sequence 17675, A
28	45.4	9.1	464	10 US-09-867-701-2947	Sequence 2947, Ap
29	42.6	8.5	929	9 US-10-093-766-14	Sequence 14, Appli
30	42.2	8.4	362	10 US-09-867-701-5411	Sequence 5411, Ap
C 31	41.8	8.3	1529	9 US-10-125-237-75	Sequence 75, Appli
C 32	41.2	8.2	81	10 US-09-864-761-24915	Sequence 24915, A
C 33	39.2	7.8	635	9 US-09-796-592-5087	Sequence 5087, Ap
C 34	39	7.8	3824	9 US-10-036-041-22	Sequence 22, Appli
35	39	7.8	3824	9 US-10-028-072-541	Sequence 541, App
36	39	7.8	3824	9 US-10-035-855-22	Sequence 22, Appli
37	39	7.8	3824	9 US-10-121-049-541	Sequence 541, App
38	39	7.8	3824	9 US-10-123-904-541	Sequence 541, App
39	39	7.8	3824	9 US-10-140-470-541	Sequence 22, Appli
40	39	7.8	3824	9 US-09-931-836-22	Sequence 541, App
41	39	7.8	3824	9 US-10-175-746-541	Sequence 541, App
42	39	7.8	3824	9 US-10-176-918-541	Sequence 541, App
43	39	7.8	3824	9 US-10-176-921-541	Sequence 541, App
44	39	7.8	3824	9 US-10-127-884-209	Sequence 209, App
45	39	7.8	3824	9 US-10-036-214-22	Sequence 22, Appli

ALIGNMENTS

RESULT 1
US-08-979-847-108
Sequence 108, Application US/08979847
Publication No. US2003003664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESIME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BRACALA, GLAUCIA
APPLICANT: KOMIRIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUXE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESS: OLIVE & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979, 847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 108:

SEQUENCE CHARACTERISTICS:
LENGTH: 1329 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-108

Query Match 85.0%; Score 426; DB 7; Length 1329;
Best Local Similarity 90.6%; Pred. No. 8,2e-136;
Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGCCAACTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGAC 61
DB 497 TGTCAAGCCAACTCCCAACAGCACTTAGGTTTCTGTGAGAGAGGAGCTGAGAGAC 556
QY 62 AGGACTAGCTGATTTCTTACGCTGACTAAGATCCCTTAAGCTTAGCTGGAAAGTGACC 121
DB 557 AGGACTAGCTGATTTCTTACGCTGACTAAGATCCCTTAAGCTTAGCTGGAAAGTGACC 616
QY 122 ACATCCACCTTTAAACACGGGGCTTGCACTTAGTTCACACCTGACCAATCAGAGCTC 181
DB 617 GATCCATCTTTAAACATGGGGCTTGCACTTAGTTCACACCTGACCAATCAGAGCTC 676
QY 182 ACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCCATCATCTATTGCTTG 241
DB 677 ACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCCATCATCTATTGCTTG 736
QY 242 AAGACACAGAGAGAGGAGCAATGATGGGATATTAACCAATCTTCAGCCGGCAAGC 301
DB 737 AAGACACAGAGAGAGGAGCAATGATGGGATATTAACCAATCTTCAGCCGGCAAGC 796
QY 302 CAACCCCTTTGGGTCCTCCCTTTGTATGGAGCTGTGTTTCAATGCTATTTCACTCT 361
DB 797 CAACCCCTTTGGGTCCTCCCTTTGTATGGAGCTGTGTTTCAATGCTATTTCACTCT 856
QY 362 ATTAATCTTGAACCTGACTCTTCTGTGCATGTTTCTTAAGGCTTGAGCTGAGCTTC 421
DB 857 ATTAATCTTGAACCTGACTCTTCTGTGCATGTTTCTTAAGGCTTGAGCTGAGCTTC 916
QY 422 GCTGGCATCCCACTGCTGTTTGGCCGACCGCAGACCCCGCTGATCCCATCCCT 481
DB 917 GCTGGCATCCCACTGCTGTTTGGCCGACCGCAGACCCCGCTGATCCCATCCCT 976
QY 482 CTGATCATGACGGGTGCTCC 501
DB 977 TTGATTCAGCAGAGTGTCC 996

RESULT 2
US-10-114-893-134
Sequence 134, Application US/10114893
Publication No. US20020193567A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M.

APPLICANT: Lavallie, Edward R.

APPLICANT: Collins-Racie, Lisa A.

APPLICANT: Evans, Cheryl

APPLICANT: Werberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Bowman, Michael R.

APPLICANT: Spaulding, Vikki

APPLICANT: Carlin-Duckett, McKeough

APPLICANT: Kelleher, Kerry S.

APPLICANT: Genetics Institute, Inc.

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

FILE REFERENCE: GI 6000-10A

CURRENT APPLICATION NUMBER: US/10/114,893

CURRENT FILING DATE: 2002-04-02

EARLIER APPLICATION NUMBER: 09/413,232

EARLIER FILING DATE: 1999-10-06

NUMBER OF SEQ ID NOS: 321

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 134
LENGTH: 2946
TYPE: DNA
ORGANISM: Homo sapiens
US-10-114-893-134

Query Match 74.9%; Score 375.4; DB 9; Length 2946;
Best Local Similarity 99.7%; Pred. No. 3e-118;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCGGCCAACTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGAC 63
DB 2553 TCGGCCAACTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGAC 2612
QY 64 GACTAGCTGATTTCTTACGCTGACTAAGATCCCTTAAGCTTAGCTGGAAAGTGACC 123
DB 2613 GACTAGCTGATTTCTTACGCTGACTAAGATCCCTTAAGCTTAGCTGGAAAGTGACC 2672
QY 124 ATCCACCTTTAAACACGGGGCTTGCACTTAGTTCACACCTGACCAATCAGAGCTCAC 183
DB 2673 ATCCACCTTTAAACACGGGGCTTGCACTTAGTTCACACCTGACCAATCAGAGCTCAC 2732
QY 184 TAAATGCTAATTAAGCAAGAGGTAAAGAAATAGCCATCATCTATTGCTTGAG 243
DB 2733 TAAATGCTAATTAAGCAAGAGGTAAAGAAATAGCCATCATCTATTGCTTGAG 2792
QY 244 AGCAGCAGAGAGGAGCAATGATGGGATATTAACCAAGCTTCGAGCCGGCAAGGCA 303
DB 2793 AGCAGCAGAGAGGAGCAATGATGGGATATTAACCAAGCTTCGAGCCGGCAAGGCA 2852
QY 304 ACCCCCTTTGGGTCCTCCCTTTGTATGGAGCTGTGTTTCAATGCTATTTCACTCTAT 363
DB 2853 ACCCCCTTTGGGTCCTCCCTTTGTATGGAGCTGTGTTTCAATGCTATTTCACTCTAT 2912
QY 364 TAAATCTTGAACCTGCA 380
DB 2913 TAAATCTTGAACCTGCA 2929

RESULT 3
US-09-902-535-1

Sequence 1, Application US/09902535
Patent No. US20020102530A1

GENERAL INFORMATION:

APPLICANT: Keish, Jr., James C.

APPLICANT: McCoy, John M.

APPLICANT: M. Sha

TITLE OF INVENTION: Methods and compositions for diagnosing

TITLE OF INVENTION: and treating preclampsia and gestational trophoblast

FILE REFERENCE: GIN-6006B4

CURRENT APPLICATION NUMBER: US/09/902,535

CURRENT FILING DATE: 2001-07-09

PRIOR APPLICATION NUMBER: 60/216,657

PRIOR FILING DATE: 2000-07-06

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 2930

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (930)...(2546)

US-09-902-535-1

Query Match 74.7%; Score 374.4; DB 10; Length 2930;
Best Local Similarity 99.7%; Pred. No. 6,6e-118;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 TCGGCCAACTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGAC 63
|||||

Db 2555 TCGGCAACTCCCAACAGCAGCTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 2614
Qy 64 GACTAGTGAATTTCTAGGCTGAGTAAAGATCCCTAAGCCTAGCTGGGAGGTGACCA 123
Db 2615 GACTAGTGAATTTCTAGGCTGAGTAAAGATCCCTAAGCCTAGCTGGGAGGTGACCA 2674
Qy 124 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGCTCAG 183
Db 2675 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGCTCAG 2734
Qy 184 TAAATGCTAATTTAGGCAAAAGCAGAGGTAAAGAAATAGCCAAATCATATTTGCTGAG 243
Db 2735 TAAATGCTAATTTAGGCAAAAGCAGAGGTAAAGAAATAGCCAAATCATATTTGCTGAG 2794
Qy 244 AGCAGCAGAGGAGGAGCATATGCGGATATATAACCAAGCTTTCGAGCCGGCAACGGCA 303
Db 2795 AGCAGCAGAGGAGGAGCATATGCGGATATATAACCAAGCTTTCGAGCCGGCAACGGCA 2854
Qy 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTCACTCTAT 363
Db 2855 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTCACTCTAT 2914
Qy 364 TAAATCTTGCACACTGC 379
Db 2915 TAAATCTTGCACACTGC 2930

RESULT 4
US-08-979-847-102
Sequence 102, Application US/08979847
Publication No. US2003003964A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESSEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIVE & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-979-847-102
Query Match 68.2%; Score 341.8; DB 12; Length 635;
Best Local Similarity 94.2%; Pred. No. 4.6e-107;
Matches 355; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 4 TCGGCAACTCCCAACAGCAGCTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 63
Db 243 TCAGCAACTCCCAACAGCAGCTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 302
Qy 64 GACTAGTGAATTTCTAGGCTGAGTAAAGATCCCTAAGCCTAGCTGGGAGGTGACCA 123
Db 303 GACTAGTGAATTTCTAGGCTGAGTAAAGATCCCTAAGCCTAGCTGGGAGGTGACCA 362
Qy 124 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGCTCAG 183
Db 363 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGCTCAG 422
Qy 184 TAAATGCTAATTTAGGCAAAAGCAGAGGTAAAGAAATAGCCAAATCATATTTGCTGAG 243
Db 423 TAAATGCTAATTTAGGCAAAAGCAGAGGTAAAGAAATAGCCAAATCATATTTGCTGAG 482
Qy 244 AGCAGCAGAGGAGGAGCATATGCGGATATATAACCAAGCTTTCGAGCCGGCAACGGCA 303
Db 483 AGCAGCAGAGGAGGAGCATATGCGGATATATAACCAAGCTTTCGAGCCGGCAACGGCA 542
Qy 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTCACTCTAT 363
Db 543 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTCACTCTAT 602
Qy 364 TAAATCTTGCACACTGC 380
Db 603 TAAATCTTGCACACTGC 619

RESULT 5
US-10-003-806-6/c
Sequence 6, Application US/10003806
Patent No. US20020119929A1
GENERAL INFORMATION:
APPLICANT: Bishop, Colin E.
APPLICANT: Agoulnik, Alexander I.
APPLICANT: Zhu, Qichao
TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
FILE REFERENCE: P02066US1/10024824
CURRENT APPLICATION NUMBER: US/10/003,806
CURRENT FILING DATE: 2001-11-02
PRIOR APPLICATION NUMBER: 60/245,872
PRIOR FILING DATE: 2000-11-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 180557
TYPE: DNA
ORGANISM: Human
US-10-003-806-6
Query Match 63.4%; Score 317.6; DB 12; Length 180557;
Best Local Similarity 85.9%; Pred. No. 1.9e-97;
Matches 395; Conservative 0; Mismatches 44; Indels 21; Gaps 3;

Qy 52 ACTGAGACAGAGACTAGCTGATTTCTTAGGCTGACCTAAGATCCCTAAGCTGCTGG 111
Db 58991 AGTGAAGACAGAGACTAGCTGATTTCTTAGGCTGACCTAAGATCCCTAAGCTGCTGG 58992
Qy 112 GAAGGTGACACATCCACTTTAAACACGGGGCTTGGCACTTAGCTCACACCTGACCAAT 171
Db 58993 GAAGGTGACCGCTTCCACTTTAAACATGGGGCTTGGCACTTAGCTCACACCGGCAAT 58992
Qy 172 C-----AGAGGCTCACTAAATGCTAATTTAGGCAAAAGCAGAGGTAAAGAAATA 222

Db 58871 CAGATAGTAAAGAGAGCTCACTAAATGCTAATTAGGCAAAAACAGAGAGTAAAGAAATA 58812
Qy 223 GCCAATCATCTATTGCTGCTGAGAGACAGACAGAGAGATGATGCGGATATATAACCCAA 282
Db 58811 GCCATCATCTATTGCTGCTGAGAGACAGACAGAGAGATGATGCGGATATATAACCCAA 58752
Qy 283 GTCCTGAGCCGCGCAACGGCAACCCCTTTGGGTCCTCCCTGCTGCTGATGAGAGCTCTGT 342
Db 58751 GCATCGAGCCAGCAACAGCTACGCTTTGGGTCCTCCCTGCTGCTGATGAGAGCTCTGT 58692
Qy 343 TTTCATGCTATTCTCTATTAATCTTGCAACTGCA--CTCTCTGCTGCATGTTCT 400
Db 58691 -----CTTCACTATTATAATCTTGCACTGCACTCTCTTTGGTCTACATTTGT 58642
Qy 401 TACGCTGAGCTGAGCTTTGCTGCTGCGCATCCACACTGCTGTTGCGCCACCGCAGAC 460
Db 58641 CATGCTGAGCTGAGCTTTCTCTGCGCATCCACACTGCTGTTGCGCGCTGCTGAGAC 58582
Qy 461 CCGCGCTGACTCCCATCCCTCTGTGATCATGCAAGGCTGTC 500
Db 58581 CTGCTGCTGACTTCCATCCGTCAGATCCGCAAGAGTGC 58542

RESULT 6
US-10-003-806-9/c

; Sequence 9, Application US/10003806
; Patent No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulink, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02066US1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-9

Query Match 63.4%; Score 317.6; DB 12; Length 180557;
Best Local Similarity 85.9%; Pred. No. 1.9e-97;
Matches 395; Conservative 0; Mismatches 44; Indels 21; Gaps 3;

Qy 52 ACTGAGAGACAGAGACTAGCTGATTTCTTAGGCTGCTAAGATCCCTAAGCTAGCTGG 111
Db 58991 ACTGAGAGACAGAGACTAGCTGATTTCTTAGGCTGCTAAGATCCCTAAGCTAGCTGG 58932
Qy 112 GAAGGTGACCACTCACTTTAACAAGGGGCTTGCACTTAGCTCACTGACCAAT 171
Db 58931 GAAGGTGACCGCTTCACTTTAACAATGGGGCTTGCACTTAGCTCACTGACCAAT 58872
Qy 172 C-----AGAGAGCTACTAAATGCTAATTAG3CAAAAGCAGGAGGTAAAGAAATA 222
Db 58871 CAGATAGTAAAGAGAGCTCACTAAATGCTAATTAG3CAAAAGCAGGAGGTAAAGAAATA 58812
Qy 223 GCCAATCATCTATTGCTGCTGAGAGACAGACAGAGAGATGATGCGGATATATAACCCAA 282
Db 58811 GCCAATCATCTATTGCTGCTGAGAGACAGACAGAGAGATGATGCGGATATATAACCCAA 58752
Qy 283 GTCCTGAGCCGCGCAACGGCAACCCCTTTGGGTCCTCCCTGCTGCTGATGAGAGCTCTGT 342
Db 58751 GCATCGAGCCAGCAACAGCTACGCTTTGGGTCCTCCCTGCTGCTGATGAGAGCTCTGT 58692
Qy 343 TTTCATGCTATTCTCTATTAATCTTGCAACTGCA--CTCTCTGCTGCATGTTCT 400
Db 58691 -----CTTCACTATTATAATCTTGCACTGCACTCTCTTTGGTCTACATTTGT 58642

Qy 401 TACGCTGAGCTGAGCTTTGCTGCTGCGCATCCACCACTGCTGTTGCCGCAACCGCAGAC 460
Db 58641 CATGCTGAGCTGAGCTTTCTCTGCGCATCCACCACTGCTGTTGCCGCTGCTGAGAC 58582
Qy 461 CCGCGCTGACTCCCATCCCTCTGTGATCATGCAAGGCTGTC 500
Db 58581 CTGCTGCTGACTTCCATCCGTCAGATCCGCAAGAGTGC 58542

RESULT 7
US-09-880-107-538/c

; Sequence 538, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-NO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 538
; LENGTH: 410
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA250958
US-09-880-107-538

Query Match 62.2%; Score 311.4; DB 10; Length 410;
Best Local Similarity 92.6%; Pred. No. 1e-96;
Matches 349; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

Qy 4 TCGGCAACCTCCCAACAGACCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
Db 377 TCAGCAACCTCCCAACAGACCTTAGGTTTCTGTTGAGAGAGGGAGCTGAGAGACAG 318
Qy 64 GACTAGCTGATTTCTGAGCTGACTAAGATCCCTAAGCTAGCTGGAAGTGACAC 123
Db 317 GACTAGCTGATTTCTGAGCTGACTAAGATCCCTAAGCTAGCTGGAAGTGACAC 258
Qy 124 ATCCACTTTAACAAGGGGCTTGCACTTAGCTCACTGACCAATCAGAGACTCAC 183
Db 257 ATCCACTTTAACAAGGGGCTTGCACTTAGCTCACTGACCAATCAGAGACTCAC 198
Qy 184 TAAATGCTAATTAGGCAAAAGAGAGGTAAAGAAATAGCAATCATCTATTGCTGTA- 242
Db 197 TAAATGCTAATTAGGCAAAAGAGAGGTAAAGAAATAGCAATCATCTATTGCTGTA- 138
Qy 243 GAGCAGACAGAGAGCAATGATGCGGATTAACCACTTTGAGCCGCGC-AAAGG 301
Db 137 GAGCAGATGAGAGAGCAATGATGCGGATTAACCACTTTGAGCCGCGCAGCAAGG 78
Qy 302 CAACCCCTTTGGGTCCTCCCTTTGATGAGAGCTGTTTTCATGCTATTCTACTCT 361
Db 77 CAACCCCTTTGGGTCCTCCCTTTGATGAGAGCTGTTTTCATGCTATTCTACTCT 18
Qy 362 ATTAATCTTGCAACTG 378
Db 17 ATTAATCTTGCAACTG 1

RESULT 8
US-09-731-231A-3/c
; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1

GENERAL INFORMATION:
APPLICANT: GUEIGER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: C1001007
CURRENT APPLICATION NUMBER: US/09/731,231A
CURRENT FILING DATE: 2000-12-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 326014
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(326014)
OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3

Query Match 60.9%; Score 305.2; DB 10; Length 326014;
Best Local Similarity 89.1%; Pred. No. 4.7e-93;
Matches 344; Conservative 0; Mismatches 33; Indels 9; Gaps 1;

Qy 4 TCGGCCAACCTCCCAACAGCAGCCTAGGTTTCTGTGAGATGGGGGAGCTGAGAGACAG 63
Db 170580 TCGGCCAACCTCCCAACAGCAGCCTAGGTTTCTGTGAGATGGGGGAGCTGAGAGACAG 170521

Qy 64 GACTGAGCTGATTTCTTCTGAGCTGATGAGATCCCTAGAGCTGAGAGAGTGAACAC 123
Db 170520 GACTGAGCTGATTTCTTCTGAGCTGATGAGATCCCTAGAGCTGAGAGAGTGAACAC 170461

Qy 124 ATCCACCTTTAAACAGGGGGCTTGAACCTTACCTACACCTGACCAATC-----AG 174
Db 170460 TTCTACCTTTAAACAGGGGGCTTGAACCTTACCTACACCTGACCAATGAGTAGAAG 170401

Qy 175 AGAGCTCACTAAATGCTATATTAGGCAAGACAGAGGTAAAGAAATAGCCATCATCTA 234
Db 170400 AGAGCTCACTAAATGCTATATTAGGCAAGACAGAGGTAAAGAAATAGCCATCATCTA 170341

Qy 235 TTGCTGAGAGACAGCAGAGGAGGACATGATCGGGATTTAAACCCAGCTTTCAGCCG 294
Db 170340 TTGCTGAGAGACAGCAGAGGAGGACATGATGCGGTATTAACCCAGGCAATTTGAGGCCA 170281

Qy 295 GGAAGGGAACCCCTTTGGGTCCTCCCTCTGTATGGGAGCTCTGTTTATGCTATT 354
Db 170280 GGAAGGGAACCCCTTTGGGTCCTCCCTCTGTATGGGAGCTCTGTTTATGCTATT 170221

Qy 355 TCACCTATTAATCTTGCACCTGCA 380
Db 170220 TCACCTATTAATCTTGCACACGCA 170195

RESULT 9
US-09-864-761-8173/c
Sequence 8173, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmiga-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 8173
LENGTH: 541
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC016663.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
US-09-864-761-8173

Query Match 57.2%; Score 286.6; DB 10; Length 541;
Best Local Similarity 82.7%; Pred. No. 3.9e-88;
Matches 383; Conservative 0; Mismatches 59; Indels 21; Gaps 4;

Qy 49 GGGACTGAGAGACAGAGCTAGCTGATTTCTAGGCTGACTAAGATCCCTAAGCTTAGC 108
Db 536 GGGACTGAGAGACAGAGCTAGCTGATTTCTAGGCTAAGATCCCTAAGCTTAGC 477

Qy 109 TGGGAAGGTGACCAATCCTTAAACAGGGGCTTGAACCTAGCTACACCTGACC 168
Db 476 TGGGAAGGTGACCAATCCTTAAACAGGGGCTTGAACCTAGCTACACCTGACC 417

Qy 169 AATC-----AGAGACTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAA 219
Db 416 AATCAGTAGTAAGACAGCTCACTAAAGGCTAATTGGGCTAAACAGAGGTAAATTA 357

Qy 220 ATAGCAATCATTTCTGAGAGACAGAGGAGGACATGATGGGATATTAACC 279
Db 356 ATAGCAATCATTTCTGAGAGACAGAGGAGGATGATGATGGGATATTAACC 297

Qy 280 CAAGCTTCGAGCGGCAAGGCAACCCCTTTGGGTCCTCCCTTTTATGGAGCTC 339
Db 296 CAAGCTTCGAGCGGCAAGGCAACCCCTTTGGGTCCTCCCTTTTATGGAGCTC 237

Qy 340 TGTTCATGCTATTTCATCTAATTAAATCTTGAACGTGACCTCT-TCGTGCTCATGTTT 398


```

QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTACCTGACCTGACCAATC-----AG 174
D 1931 ATCCATCTTTTAAACATGGGGCTTGCACTTACCTGACCAATCAGGTATTTAAAG 1872
QY 175 AGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATGCAATCATCTA 234
D 1871 AGAGCTCACTAAATGCTAATTAGGCAAGAAACAGAGGTAAACAAATGCAATCATCTA 1812
QY 235 TTGCTGAGACACAGACAGAGGAGCAATGATCGGATATTAACCCAACTCTTGAGCCG 294
D 1811 TCGCTGAGACACAGAGGAGGAGCAATGATCGGATATTAACCCAACTCTTGAGCTG 1752
QY 295 GCAACGGCAACCCCTTTGGGTCCCTCTTTGATGAGAGCTGTGTTTCAGCTAAT 354
D 1751 GCAACAGCTACCTCTTTGGGTCCCTCTTTGATGAGAGCTGTGTTTCAGCTACT 1692

```

```

RESULT 12
US-09-764-869-1654/c
; Sequence 1654, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1654
; LENGTH: 15425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-869-1654

```

```

Query Match 53.3%; Score 266.8; DB 10; Length 15425;
Best Local Similarity 88.1%; Pred. No. 1.4e-80;
Matches 317; Conservative 0; Mismatches 32; Indels 11; Gaps 2;

```

```

QY 4 TCGGCAACCTTCCCAACAGCACTTAGTTTCTGTTGAGATGGGGACGTGAGACAG 63
D 2049 TTGGGCAACCTCCCAACAGCACTTAGTTTCTGTTGAGATGGGGAGCT--GAGACAG 1992
QY 64 GACTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGAGAGTGACAC 123
D 1991 AACTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGAGAGTGACTGC 1932
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTACCTGACCAATC-----AG 174
D 1931 ATCCATCTTTTAAACATGGGGCTTGCACTTACCTGACCAATCAGGTATTTAAAG 1872
QY 175 AGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATGCAATCATCTA 234
D 1871 AGAGCTCACTAAATGCTAATTAGGCAAGAAACAGAGGTAAACAAATGCAATCATCTA 1812
QY 235 TTGCTGAGACACAGACAGGAGCAATGATCGGATATTAACCCAACTCTTGAGCCG 294
D 1811 TCGCTGAGACACAGAGGAGGAGCAATGATCGGATATTAACCCAACTCTTGAGCTG 1752
QY 295 GCAACGGCAACCCCTTTGGGTCCCTCTTTGATGAGAGCTGTGTTTCAGCTAAT 354
D 1751 GCAACAGCTACCTCTTTGGGTCCCTCTTTGATGAGAGCTGTGTTTCAGCTACT 1692

```

```

RESULT 13
US-10-040-916-50
; Sequence 50, Application US/10040916
; Patent No. US20020146769A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John
; Lavallee, Edward
;

```

```

; Racie, Lisa
; Merberg, David
; Treacy, Maurice
; Evans, Cheryl
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/040,916
; FILING DATE: 07-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/887,029
; FILING DATE: 07-FEB-1997
; APPLICATION NUMBER: 08/686,878
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-040-916-50

```

```

Query Match 47.6%; Score 238.4; DB 12; Length 279;
Best Local Similarity 92.4%; Pred. No. 9.9e-72;
Matches 242; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

```

```

QY 119 ACCACATCCACCTTTAAACAGGGGGCTTGCACTTACCTGACCACTGACAGAG 178
D 1 RCACATCCACCTTTAAACAGGGGNTTGCAAAAAGATNACATTTGACCAATCAGAGAG 60
QY 179 CTCATTAATGCTAATTAGGCAAGACAGAGGTAAAGAAATGCAATCATCTATTGC 238
D 61 NTCANTAAATGATNATTNGCCAAACAGAGGTAAAGAAATGCCAATCATCTATTGC 120
QY 239 CTGAGAGACAGCAGAGGAGCAATGATCGGATATTAACCCAACTCTTGAGCCGCA 298
D 121 CTGAGAGACAGCAGAGGAGCAATGATCGGATATTAACCCAACTCTTGAGCCGCA 180
QY 299 GGGCAACCCCTTTGGGTCCCTCTTTGATGAGAGCTGTGTTTCAGCTATTTCAC 358
D 181 GGGCAACCCCTTTGGGTCCCTCTTTGATGAGAGCTGTGTTTCAGCTATTTCAN 240
QY 359 TCTATTAATCTTGCAACTGCA 380
D 241 TTTATTAATTTGCAACTGCA 262

```

```

RESULT 14
US-09-864-761-14951/c
; Sequence 14951, Application US/09864761
;

```

```
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14951
LENGTH: 569
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP00233.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
US-09-864-761-14951

Query Match 40.7%; Score 203.8; DB 10; Length 569;
Best Local Similarity 79.3%; Pred. No. 1,1e-59;
Matches 337; Conservative 0; Mismatches 62; Indels 26; Gaps 7;
```

```
Qy 171 TC-----AGAGCTCACTAAATGCTAATTAGC-AAAGACAGAGCTAAAGAA 220
Db 295 TCAGTAGTAAGAGGGTTCACTGAATTAACAATTAGCTAAAGCAGAGGTAAAGAAA 236
Qy 221 TAG-CCATCATCTATTCCTGAGAGCAGACAGAGGACAAATGATCGGATTAACC 279
Db 235 TAGCAAAATCATATATGCTTAACAGACAGGGGAGGACATGATGGGATTAAC 176
Qy 280 CAA--GTCTTGAGCCCGGCAACGCCCCCTTTGGGTCCCTCCCTTGATGGAGC 337
Db 175 CCAGGATTTAAACCCGGAGATGGCAACCCCTTTGGGTCCCTCCCATGTATGGAGC 116
Qy 338 TCTGTTTCATGCTATTTCACTCTATTAATCTTGCAATG--CACTCTGCTCATG 395
Db 115 TCTGT-----TTCACTCTGTAAATCTTGCAACTGTAACTCTCTGTGCTAGT 66
Qy 396 TTCTTTCAGGCTTGAAGCTTTCGCTGCGCATCCACCACTGCTTTGGCCGCCACG 455
Db 65 TTGTCGGCTCAAGCTGAACCTTTTGTCTACCTCTTACCACTGCTGTTCCCGGTAG 6
Qy 456 CAGAC 460
Db 5 CAGAC 1
```

```
RESULT 15
US-09-864-761-20462/c
Sequence 20462, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
```

```

? PRIOR APPLICATION NUMBER: US 09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1.1
? SEQ ID NO 20462
? LENGTH: 246

```

Query Match	29.2%	Score	146.2	DB	10	Length	246
Best Local Similarity	88.5%	Pred. No.	4.4e-40				
Matches	185	Conservative	0	Mismatches	13	Indels	11
						Gaps	2

Qy	232	CTATTGCGCTGAGACACACAGCAGAGAGGGACATGATCGGGAATATTAACCCCAAGCTTCGAG	291
Db	246	CTGTTCGCTGAGAGACACAGCGGGAGGGACATTTATCAGAGATTTAAACCCCAAGCATTCGAG	187
Qy	292	CCGGCAACGCGCAACCCCTTTTGAGTCCCTCCCTTTGATGAGGAGCTCTGTTCATGCT	351
Db	186	CTGGCAACGGTAACCCCTCTTTGGTGCCCTCTCTTGATGAGGAGCTCTAT-----	136
Qy	352	ATTTCACCTCTATTTAAATTTTGCAACTGCACTCTTCGTGCATGTTCTTAAGGCTTGAG	411
Db	135	CTTTCACCTCTATTTAAATTTTGCAACTGTACTCTTCGTGCATGTTGTTGTTAAC-GCTTGAG	78
Qy	412	CTGAGCTTTGCTGCGCCATCCACCACTGC	440
Db	77	CTGAGCTTTGCTGCGCCATCCACCACTGC	49

Search completed: April 17, 2003, 07:49:41
Job time : 300.415 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:30:25 ; Search time 1268.84 Seconds

(without alignments)
6394.748 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501
Sequence: 1 gctcgcgcacactcccca.....ctgcatcatgcaggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estcov:*
6: em_esttpl:*
7: em_esttro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	376	75.0	440	12	BE732673 601571305
2	372.8	74.4	494	9	AA781423 aj26c03.8
3	372.2	74.3	436	9	A1128526 qc61h10.x
4	365.8	73.0	415	9	A1128496 qc61e08.x
5	360.4	71.9	388	14	H01325 y199e01.s1
6	355.4	70.9	727	17	AG029908 Pan trogl

C	7	348.4	69.5	422	14	NS3177	NS3177 yv56h11.s1
	8	345.2	68.9	771	13	BI087886	BI087886 602852690
	9	343.8	68.6	683	17	AG102951	AG102951 Pan trogl
C	10	339.6	67.8	385	14	NS5091	NS5091 yv43e03.s1
	11	337.8	67.4	458	14	R76086	R76086 y171b03.s1
C	12	332.4	66.3	522	10	AW971553	AW971553 EST383642
	13	329	65.7	363	14	T69704	T69704 yd13a03.s1
C	14	326.6	65.2	342	9	AA860368	AA860368 aj59c05.s
C	15	326.4	65.1	424	14	R27412	R27412 yh46d11.s1
C	16	324.4	64.8	342	14	T47345	T47345 yb10h02.s1
C	17	318.8	63.6	438	14	R77278	R77278 y175d06.s1
C	18	317.6	63.4	443	9	AA837267	AA837267 od26b10.s
C	19	312.6	62.4	433	9	A1379210	A1379210 c601g11.x
C	20	312.6	62.4	490	9	A1598135	A1598135 tml4a10.x
C	21	311.4	62.2	410	9	AA250958	AA250958 zso7d10.s
C	22	311	62.1	425	9	A1570707	A1570707 tm79g09.x
C	23	309.4	61.8	446	9	A1393478	A1393478 tg45g04.x
C	24	307.2	61.3	921	17	BH149565	BH149565 ENTQ48TR
C	25	305.4	61.0	463	14	R68685	R68685 y14g06.s1
C	26	302.6	60.4	404	14	R27389	R27389 yh46a09.s1
C	27	296.2	59.1	328	9	AA729556	AA729556 nx58c05.s
C	28	293.6	58.6	485	10	AW511366	AW511366 hd45h03.x
C	29	293	58.5	431	9	AA552941	AA552941 nk61a10.s
C	30	290.8	58.0	471	9	AA709471	AA709471 zf19h06.s
C	31	289.6	57.8	653	17	AG033781	AG033781 Pan trogl
C	32	287.8	57.4	701	17	AG126669	AG126669 Pan trogl
C	33	285	56.9	611	17	AQ381711	AQ381711 RPCI11-16
C	34	284	56.7	712	17	AO892947	AO892947 HS 3131.B
C	35	281.2	56.1	722	17	AG049481	AG049481 Pan trogl
	36	279.2	55.7	503	14	BM723292	BM723292 UI-E-EUO-
C	37	278.4	55.6	530	14	BM682297	BM682297 UI-E-EUO-
C	38	277.6	55.4	470	9	A1074704	A1074704 ox83d05.s
C	39	277.6	55.1	446	9	A1288235	A1288235 q18a08.x
C	40	276.2	55.1	960	17	AO900343	AO900343 HS 3179.B
	41	273.2	54.5	893	14	BQ437925	BQ437925 AGENCOURT
	42	272.8	54.5	609	17	AG066901	AG066901 Pan trogl
C	43	272.4	54.4	326	14	D29167	D29167 HUMNR203 Hu
	44	272.4	54.4	447	12	BF919416	BF919416 QVO-NT015
	45	271.8	54.3	681	10	AV722664	AV722664 AV722664

ALIGNMENTS

RESULT 1
BE732673
LOCUS
DEFINITION
601571305F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3925728 5',
mRNA sequence.
ACCESSION
BE732673
VERSION
BE732673.1 GI:10146665
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 440)
NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
CONTACT
Contact: Robert Strauberg, Ph.D.
COMMENT
Email: cgabds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
Plate: L1CM752 row: j column: 01
High quality sequence, stop: 440.
Location/Qualifiers
1..440
/organism="Homo sapiens"

FEATURES

source

/db_xref="taxon:9606"
/clone="IMAGE:3925728"
/clone_lib="NIH_MGC_21"
/issue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Placenta; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: 3GCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and SuperScript II RT (Life Technologies)."

BASE COUNT 124 a 117 c 101 g 98 t
ORIGIN

Query Match 75.0%; Score 376; DB 12; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.8e-111;
Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TCGGCCAACCTCCCAACAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 65 TCGGCCAACCTCCCAACAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 124
OY 64 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTACAC 123
DB 125 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTACAC 184
OY 124 ATCCACCTTTAAACACGGGGCTTGGCACTTACTACACCTTACCAATCAGAGACTCAG 183
DB 185 ATCCACCTTTAAACACGGGGCTTGGCACTTACTACACCTTACCAATCAGAGACTCAG 244
OY 184 TAAATGCTAATTAGCAAGAGAGAGTAAAGTAATGCAATCATCTATTGCTTGA 243
DB 245 TAAATGCTAATTAGCAAGAGAGAGTAAAGTAATGCAATCATCTATTGCTTGA 304
OY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGCTTTCGAGCGGCAACGCA 303
DB 305 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGCTTTCGAGCGGCAACGCA 364
OY 304 ACCCCCTTTGGGTCCCTTCTTGTATGGAGCTTGTTCATGCTATTTCACCTCAT 363
DB 365 ACCCCCTTTGGGTCCCTTCTTGTATGGAGCTTGTTCATGCTATTTCACCTCAT 424
OY 364 TAAATCTTGCACTGC 379
DB 425 TAAATCTTGCACTGC 440

RESULT 2
AA781423 494 bp mRNA linear EST 31-DEC-1998
LOCUS aJ36C03.81 Soares_testis_NHT Homo sapiens cDNA clone 1391428 3'
DEFINITION similar to contains PTR7.1 PTR7 repetitive element; mRNA
sequence.

ACCESSION AA781423

VERSION AA781423.1 GI:2840754

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rcmail.nih.gov
cDNA Library Preparation: M. Benito Soares, Ph.D., M. Fatima Bonaldo
, Ph.D. Library Arrayed by: Greg Lemmon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnl.gov/bbrp/image/image.html
Insert Length: 1645 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 475.
Location/Qualifiers

FEATURES
source

1. 494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="1391428"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTAGAGTGGAGGCGGCCCAATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Benito Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 111 c 133 g 141 t 1 others

Query Match 74.4%; Score 372.8; DB 9; Length 494;
Best Local Similarity 99.2%; Pred. No. 2.1e-110;
Matches 374; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 TCGGCCAACCTCCCAACAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 382 TCGGCCAACCTCCCAACAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 323
OY 64 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTACAC 123
DB 322 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTACAC 263
OY 124 ATCCACCTTTAAACACGGGGCTTGGCACTTACTACACCTTACCAATCAGAGACTCAG 183
DB 262 ATCCACCTTTAAACACGGGGCTTGGCACTTACTACACCTTACCAATCAGAGACTCAG 203
OY 262 TAAATGCTAATTAGCAAGAGAGTAAAGTAATGCAATCATCTATTGCTTGA 243
DB 202 TAAATGCTAATTAGCAAGAGAGTAAAGTAATGCAATCATCTATTGCTTGA 143
OY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGCTTTCGAGCGGCAACGCA 303
DB 142 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGCTTTCGAGCGGCAACGCA 83
OY 304 ACCCCCTTTGGGTCCCTTCTTGTATGGAGCTTGTTCATGCTATTTCACCTCAT 363
DB 82 ACCCCCTTTGGGTCCCTTCTTGTATGGAGCTTGTTCATGCTATTTCACCTCAT 23
OY 364 TAAATCTTGCACTGC 380
DB 22 TAAATCTTGCACTGC 6

RESULT 3
A1128526 436 bp mRNA linear EST 27-OCT-1998
LOCUS A1128526

DEFINITION g66h10.x1 Soares Placenta 8c05week 2bDH8bc09M Homo sapiens cDNA
clone IMAGE:1724147 3' similar to contains PTR7.1 PTR5 repetitive
element; mRNA sequence.

ACCESSION A1128526

VERSION A1128526.1 GI:3597040

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 436)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 720 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence, stop: 428.
Location/Qualifiers
1. 436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_placenta_8to9weeks_2NBH8to9W"
/dev_stage="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Patricia Bernaldo."

BASE COUNT 95 a 96 c 106 g 139 t
ORIGIN
Query Match 74.3%; Score 372.2; DB 9; Length 436;
Best Local Similarity 99.2%; Pred. No. 3.1e-110;
Matches 374; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TCGGCAACTCCCAAGACGACTAGGTTTCTGTGAGATGGGGAGTGAAGAG 63
DB 399 TCGGCAACTCCCAAGACGACTAGGTTTCTGTGAGATGGGGAGTGAAGAG 340
QY 64 GACTAGCTGATTTCTTGAAGTGAATCCCTAAGACCTAGTGGAGGTGACAC 123
DB 339 GACTAGCTGATTTCTTGAAGTGAATCCCTAAGACCTAGTGGAGGTGACAC 280
QY 124 ATCCACCTTTAAACGCGGGCTTGCACTTACCTACCACTTACCACTACAGAGACTC 183
DB 279 ATCCACCTTTAAACGCGGGCTTGCACTTACCTACCACTTACCACTACAGAGACTC 220
QY 184 TAAATGCTAATTAGCAAGAAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 243
DB 219 TAAATGCTAATTAGCAAGAAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 160
QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCA 303
DB 159 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCA 100
QY 304 ACCCTTTGGGTCCCTTCTTGTATGGAGCTGTGTTTCAATGCTATTTCACTCTAT 363
DB 99 ACCCTTTGGGTCCCTTCTTGTATGGAGCTGTGTTTCAATGCTATTTCACTCTAT 40
QY 364 TAAATCTTGCAACTGCA 380
DB 39 TAAATCTTGCAACTGCA 23

RESULT 4
LOCUS A1128496 415 bp mRNA linear EST 27-OCT-1998
DEFINITION gct1e08.x1 Soares_placenta_8to9weeks_2NBH8to9W Homo sapiens cDNA
clone IMAGE:1714118 3' similar to contains PTR.tl PTR7 repetitive
element // mRNA sequence.
ACCESSION A1128496

VERSION A1128496.1 GI:3597010
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1184 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence, stop: 413.
Location/Qualifiers
1. 415
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_placenta_8to9weeks_2NBH8to9W"
/dev_stage="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Patricia Bernaldo."

BASE COUNT 94 a 96 c 107 g 116 t 2 others
ORIGIN
Query Match 73.0%; Score 365.8; DB 9; Length 415;
Best Local Similarity 97.6%; Pred. No. 3.7e-108;
Matches 370; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GTCGCGCAACTCCCAAGACGACTAGGTTTCTGTGAGATGGGGAGTGAAGAG 60
DB 379 GTCGCGCGCAACTCCCAAGACGACTAGGTTTCTGTGAGATGGGGAGTGAAGAG 320
QY 61 CAGACTAGCTGATTTCTTGAAGTGAATCCCTAAGACCTAGTGGAGGTGAC 120
DB 319 CAGACTAGCTGATTTCTTGAAGTGAATCCCTAAGACCTAGTGGAGGTGAC 260
QY 121 CACATCCACCTTTAAACGCGGGCTTGCACTTACCTACCACTTACCACTACAGAGACT 180
DB 259 CACATCCACCTTTAAACGCGGGCTTGCACTTACCTACCACTTACCACTACAGAGACT 200
QY 181 CACTAATGCTAATTAGCAAGAAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 240
DB 199 CACTAATGCTAATTAGCAAGAAGAGAGTAAAGAAATAGCAATCATCTATTGCTGAG 140
QY 241 GAGAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACG 300
DB 139 GAGAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACG 80
QY 301 GCAACCCCTTTGGGTCCCTTCTTGTATGGAGCTGTGTTTCAATGCTATTTCACTCTAT 360
DB 79 GCAACCCCTTTGGGTCCCTTCTTGTATGGAGCTGTGTTTCAATGCTATTTCACTCT 20
QY 361 TATTAAATCTTGCAACTGC 379
DB 19 TATTAAATCTTGCAACTGC 1

RESULT 5

[illegible]

QY	244	AGCAGAGAGAGAGGACAAATGATGGGATATTAATTAACCCAAAGCTTCGAGCCGGCAACGGCA	303
Db	137	AGCAGAGAGAGAGGACAAATATGCGGATATTAACCCAAAGCTTCGAGCCGGCAACGGCA	78
QY	304	ACCCCTTTGGGTCCCTCCCTTGTATAGGGAGAGCTGTGTTTCATGCTATTTCACTCTAT	363
Db	77	AGCCCTTTGGGTCCCTCCCTTGTATAGGGAGAGCTGTGTTTCATGCTATTTCACTCTAT	18
QY	364	TAAATCTTGCAA	375
Db	17	TAAATCTTGCAA	6
RESULT 6			
AG023908			
LOCUS	AG029908	727 bp	DNA linear GSS 01-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-002C04.R, genomic survey sequence.		
ACCESSION	AG029908		
VERSION	AG029908.1	GI:16556781	
KEYWORDS	GSS.		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male		
ORGANISM	Pan troglodytes		
REFERENCE	BAC library clone:PTB-002C04.R.		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
REFERENCE	1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.		
AUTHORS	2 (bases 1 to 727)		
JOURNAL	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.		
COMMENT	Direct Submission		
	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suichi-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.		
	PRIMERS		
	Sequencing: MJ3Rev		
	LIBRARY		
	Vector : pKS145		
	R.Site 1 : SacI		
	R.Site 2 : SacI.		
FEATURES	Location/Qualifiers		
source	1..727		
	/organism="Pan troglodytes"		
	/db_xref="taxon:9598"		
	/clone="PTB-002C04.R"		
	/sex="male"		
	/cell_type="lymphoblast"		
	/clone_lib="PTB Chimpanzee Male BAC Library"		
BASE COUNT	170 a 215 c 175 g 166 t	1	others
ORIGIN			
Query Match	70.9%; Score 355.4; DB 17; Length 727;		
Best Local Similarity	91.1%; Pred. No. 1.2e-104;		
Matches	411; Conservative 0; Mismatches 36; Indels 4; Gaps 3;		
QY	52	ACTGAGACAGAGATTAAGTGGATTTTCTAGGCTACATTAAGATTCCTTAAGCTGACGCG	111
Db	275	ATTGAGAGACAGAGATTAAGTGGATTTTCTAGGCTGACATTAAGATTCCTTAAGCTGACGCG	334
QY	112	GAGGTGACCAATCATCCTTTAAACAGGGGCTTGCAACTGACTGCAACCTGCAAT	171
Db	335	GAGGTGACGAGATCATCCTTTAAACAGGGGCTTGCAACTTGTCTACACCCAACTAT	394
QY	172	CAGAGAGCTCACTAATATGCTATTATTAGGCMAAGACAGAGGCTAAAGAAATAGCCANTAT	231

	Db	395	CGGAGAGCTCACTTAAATGGCTAATTAGGCATAAACAAGAGGTGAAGAATACCATTAAT	454
	Oy	232	CTATTGCCGTGAGACAGCAGCAGAGGGA CAATGATCGGATATTAACCAGTCCTTGAG	291
	Db	455	CTATTGCCGTGAGACAGCAGTGGGAAGGATTAAGGATCGGGA TTTAAACCAAGCATTCAAG	514
	Oy	282	CCGGCACAGGCAACCCCTTTGGGTCCCCTCTTTGTA TGGAGACTCTGTTTTATGCT	351
	Db	515	CCAGCAACGGCAACCCCTTTGGGT - CCGTCCCTCTGTA TGGAGACTCTGTTTTACACT	573
	Oy	352	ATTTCACCTTAATATCTTGCA CTGCACTCTTCGTGGCA TGTTCTTCAAGGCTTAG	411
	Db	574	ATTTCACCTTAATATCTTGCACTTGCACTTTCTGGTC GGTGTTTGTAGAGGCTCAG	633
	Oy	412	CTGAGCTTTCGCTCGGCATCCACCACTGCTGTT - TGCCGCAACCCAGACCCCGCTGA	470
	Db	634	CTGAGCTTTCCTTCGCCCATCCACCACTGCTGTTNTG CGCGCCGTCCGACACCCCGCTGA	693
	Oy	471	CTCCCATCTCTCTGGATCATGCAAGGCTGTC	501
	Db	694	CTCCCATCTCTCT - GATCCAGCAGAGGTGTC	722
RESULT 7				
N53177/c	LOCUS			
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
TITLE				
JOURNAL				
MEDLINE				
COMMENT				
FEATURES				
Source				

BASE COUNT	95 a	95 c	111 g	120 t	1 others
ORIGIN					
Query Match	69.5%	Score 348.4;	DB 14;	Length 422;	
Best Local Similarity	98.7%	Pred. No. 1.be-102;			
Matches 372;	Conservative	0;	Mismatches 2;	Indels 3;	Gaps 2;
OY	4	TCGGCAACCTCCCCAAGACGACTTAGGTTTCCGTGTGAGATGGGGAGACTGAGAGACAG	63		
Db	375	TCGGCAACCTCCCCAAGACGACTTAGGTTTCCGTGTGAGATGGGGAGACTGAGAGACAG	316		
OY	64	GACTAGCTGATTTCTTAGCTGACTAAGAAATCCCTAAGCTAGCTGGGAAGGTGACCAC	123		
Db	315	GACTAGCTGATTTCTTAGCTGACTAAGAAATCCCTAAGCTAGCTGGGAAGGTGACCAC	256		
OY	124	ATCCACTTTAAACACGGGGCTTGCACTTAGCTACACCTGACCAATCAGAGAGCTCAC	183		
Db	255	ATCCACTTTAAACACGGGGCTTGCACTTAGCTACACCTGACCAATCAGAGAGCTCAC	196		
OY	184	TAAATGCTATTAGGCAACAGACGGAGGTAAAGAAATAGCCATCATCTATTGCTGTAG	243		
Db	195	TAAATGCTATTAGGCAACAGACGGAGGTAAAGAAATAGCCATCATCTATTGCTGTAG	136		
OY	244	AGCACAGCAGAGGAGACATGATCGGGATATTAACCCAGCTTTCGAGCCGGCAACGGCA	303		
Db	135	AGCACAGCAGAGGAGACATGATCGGGATATTAACCCAGCTTTCGAGCCGGCAACGGCA	77		
OY	304	ACCCCTTTGGGTCCCTCCTCTTTGATGGAGCTCTGTTTCACTGCTATTACTCTAT	363		
Db	76	ACCCCTTTGGGTCCCTCCTCTTTGATGGAGCTCTGTTTCACTGCTATTACTCTAT	19		
OY	364	TAAATCTGCAACTGCA	380		
Db	18	TAAATCTGCAACTGCA	2		
RESULT 8					
LOCUS	BI087886	771 bp	mRNA	linear	EST 20-JUN-2001
DEFINITION	6028526390F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4993894 5',				
ACCESSION	BI087886				
VERSION	BI087886.1	GI:14506216			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	NIH-MGC http://mgc.ncl.nih.gov/				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgabbs-remail.nih.gov				
	Tissue Procurement: ATCC				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: Incyte Genomics, Inc.				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LINL at:				
	http://image.llnl.gov				
	Plate: L1AM1015 row: d column: 23				
FEATURES	High quality sequence, stop: 762.				
source	Location/Qualifiers				
	1..771				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:4993894"				
	/clone_lib="NIH_MGC_10"				

```
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: PCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
BASE COUNT      225 a      206 c      166 g      174 t
ORIGIN
Query Match      68.9%; Score 345.2; DB 13; Length 771;
Best Local Similarity 97.4%; Pred. No. 2.6e-101;
Matches 372; Conservative 0; Mismatches 8; Indels 2; Gaps 2;
QY 1 GTCTGGCCCAACCTCCCAACAGCACTTAGTCTTCTGT-TGAGATGGGGAGCTGAGAG 59
Db 377 GTCTGGCCCAACCTCCCAACAGCACTTAGTCTTCTGTGAGATGGGGAGCTGAGAG 436
QY 60 ACAGACTAGCTGAGATTTCTTCTGAGCTGAGTAACTCCCTAAGCTTGGAGAGTGA 119
Db 437 ACAGACTAGCTGAGATTTCTTCTGAGCTGAGTAACTCCCTAAGCTTGGAGAGTGA 496
QY 120 CCACATCCACTTTAAACAAGGGGCTTGCACTTAGCTCACACTGACCAATCAGAGAGC 179
Db 497 CCACATCCACTTTAAACAAGGGGCTTGCACTTAGCTCACACTGACCAATCAGAGAGC 556
QY 180 TCCTAAATGCTAAATTAGGCAAGAGAGGTA-AGAAATAGCCAAATCAATCTATTGC 238
Db 557 TCCTAAATGCTAAATTAGGCAAGAGAGGTA-AGAAATAGCCAAATCAATCTATTGC 616
QY 239 CTGAGAGCAGCAGAGAGAGGAGCAATGATGGATATTAACCAAGTCTTCCAGCCGCAA 298
Db 617 CTGAGAGCAGCAGAGAGAGGAGCAATGATGGATATTAACCAAGTCTTCCAGCCGCAA 676
QY 299 CGGCAACCCCTTTGGGTCCTCCCTTGTATG-3GAGCTGTGTTTCATGCTATTTCAC 358
Db 677 CGGCAACCCCTTTGGGTCCTCCCTTGTATG-3GAGCTGTGTTTCATGCTATTTCAC 736
QY 359 TCTATTAAATCTTGCAACTGCA 380
Db 737 TCTATTAAATCTTGCAACTGCA 758

RESULT 9
AG102951      683 bp      DNA      linear      GSS 03-NOV-2001
LOCUS      Pan troglodytes DNA, clone: PTB-106G16.F, genomic survey sequence.
DEFINITION      AG102951
ACCESSION      AG102951.1 GI:16723468
VERSION
KEYWORDS
SOURCE
ORGANISM
Pan troglodytes male lymphoblast DNA, clone lib: PTB Chimpanzee Male
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE
1
AUTHORS      Fujiyama, A., Hatori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoki, Y., Matanabe, H. and Sakaki, Y.
TITLE      BAC end sequences of library PTB
JOURNAL
Unpublished
2 (bases 1 to 683)
Fujiyama, A., Hatori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoki, Y., Matanabe, H. and Sakaki, Y.
DIRECT SUBMISSION
Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-1-22 Suohiro-cho, Tsukuba, Ibaraki, 305-8565, Japan
E-mail: chipmbe@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the Red process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
```

```
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. 683
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-106G16.F"
/sex="male"
/cell_type="lymphoblast"
/collection="PTB Chimpanzee Male BAC Library"
BASE COUNT      162 a      186 c      167 g      167 t
ORIGIN
Query Match      68.8%; Score 343.8; DB 17; Length 683;
Best Local Similarity 89.8%; Pred. No. 7.1e-101;
Matches 403; Conservative 0; Mismatches 42; Indels 4; Gaps 3;
QY 52 ACTGAGACAGAGACTAGCTGATTTCTTAGGCTGACTGACTAAGAAATCCCTAAGCTAGCTGG 111
Db 108 ATTGAGAGAAAGAGCTAGCTGATTTCTTAGGCTGACTGACTAAGAAATCCCTAAGCTAGCTGG 167
QY 112 GAAGTGACCATTCACCTTTAAACAAGGGGCTTGCAACTTACTCAGACCTAGCAAT 171
Db 168 GAAGTGACCATTCACCTTTAAACAAGGGGCTTGCAACTTACTCAGACCTAGCAAT 227
QY 172 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAGAGGTAAGAAATAGCAATCAT 231
Db 228 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAGAGGTAAGAAATAGCAATCAT 286
QY 232 CTATTGCTGAGACAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAG 291
Db 287 CTATTGCTGAGACAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAG 346
QY 292 CCGGCAAGGCAACCCCTTTGGGTCCTCCCTTGTATGAGAGCTCTGTTCATGCT 351
Db 347 CCGGCAAGGCAACCCCTTTGGGTCCTCCCTTGTATGAGAGCTCTGTTCATGCT 404
QY 352 ATTTCACCTAATTAATCTTCACTGACACTCTCTGATGCTATGTTCTTACGCTTGA 411
Db 405 ATTTCACCTAATTAATCTTCACTGACACTCTCTGATGCTATGTTCTTACGCTTGA 464
QY 412 CTGAGCTTTCGCTCGCATCCACCACTGCTG-TTGGCCGCAACCGCAGACCCGCGCTGA 470
Db 465 CTGAGCTTTCGCTCGCTGCTGCTGCTGCTGTTGTTGTCATGTCAGACCCAGCTGTA 524
QY 471 CTCCCATCCCTCTGATGATCAGAGTGT 499
Db 525 CTCCCATCCCTCTGATGATCAGAGTGT 553

RESULT 10
NS5091/c
LOCUS      NS5091      385 bp      mRNA      linear      EST 28-JAN-1997
DEFINITION      yv43e03.81 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:245500 3', mRNA sequence.
ACCESSION      NS5091
VERSION      NS5091.1 GI:1197970
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 385)
Haller, L., Lemon, G., Becker, M., Bonaldo, M.F., Chippelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W., Hawkins
, B., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore
, B., Morris, M., Parsons, J., Prange, C., Riekin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaekis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
```


QY	304	ACCCCTTTGGGTCCTCCCTTTGATGGAG-TCTGTTTCATGCTATTCACCTAT	363
Db	59	A-CCCCCTTTGGGTCCTCCCTTTGATGGAG-TCTGTTTCATGCTATTCACCTAT	1
RESULT 12			
LOCUS	AM971553/c	522 bp	mRNA linear EST 01-JUN-2000
DEFINITION	EST183642	MAGE resequences, MAGI, Homo sapiens	CDNA, mRNA sequence.
ACCESSION	AM971553		
VERSION	AM971553.1	GI:8161399	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Hedde, P., Qi, R., Abernathy, K., Daarap, S., Gaspard, R., Gay, C., Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and Quackenbush, J.		
TITLE	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element CDNA microarray		
COMMENT	Unpublished (2000)		
JOURNAL	Contact: John Quackenbush		
	The Institute for Genomic Research		
	9712 Medical Center Dr., Rockville, MD 20850, USA		
	Tel: 301 838 3528		
	Fax: 301 838 0208		
	Email: john@igr.org		
	Plate: 292		
	Seq primer: Forward.		
FEATURES	Location/Qualifiers		
source	1..522		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="MAGE resequences, MAGI"		
	/note="Vector: pBluescriptSKm"		
BASE COUNT	125 a 115 c 134 g 143 t		
ORIGIN			
Query Match	66.3%; Score 332.4; DB 10; Length 522;		
Best Local Similarity	93.0%; Pred. No. 3.3e-97;		
Matches 348; Conservative	0; Mismatches 26; Indels 0; Gaps 0.		
QY	4	TCGGCCAACTCCCAACAGCACTTAGTTCTCTTTGATGGGGAAGCTGAGACAG	63
Db	374	TCAGCAACTCACAACAGCACTTGCTGTTCTTTGAGAGGAGGAGCTGAGACAG	315
QY	64	GACTAGCTGATTTCTTAGCTGACTAAGATCCCTAAGCTTAGCTGGAGAGTGCAC	123
Db	314	GACTAGCTGATTTCTTAGCGGATTAAGATCCCTAAGCTTAGCTGGAGAGTGCAC	255
QY	124	ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAATGAGAGCTCAC	183
Db	254	ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAATGAGAGCTCAC	195
QY	184	TAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCCATCATATTGCTCGAG	243
Db	194	TAAATGCTAATTTAGGCAAGAAACAGAGGTAAAGAAATAGCCATCATATTGCTCGAG	135
QY	244	AGCACAGCAGAGGAGCAATGATCGGATATTAACCTAAGCTTTCAGCGGACAGCGCA	303
Db	134	AGCACAGTGGAGGAGCAAGATTTGCAATATAAACCAAGCATTTGAGCAGACGAGCA	75
QY	304	ACCCCTTTGGGTCCTCCCTTTGATGGAGCTCTGTTTCATGCTATTCACCTAT	363
Db	74	ACCCCTTTGGGTCCTCCCTTTGATGGAGCTCTGTTTCATGCTATTCACCTAT	15
QY	364	TAAATCTGCACT 377	
Db	14	TAAATCTGCACT 1	

	RESULT	13
T69704/c	LOCUS	
YJ33A03.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:66988 3', mRNA sequence.	DEFINITION	
T69704	ACCESSION	
T69704.1 GI:680852	VERSION	
EST .	KEYWORDS	
human.	SOURCE	
Homo sapiens	ORGANISM	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	REFERENCE	
1 (bases 1 to 363)	AUTHORS	
Hillier L., Clark N., Dubque T., Elliston K., Hawkins M., Holman M., Hulman M., Kucaba T., Le M., Lennon G., Merra M., Parsons J., Rifkin, L.L., Rohlfing T., Soares M., Tan F., Trevaskis E., Waterston R., Williamson A., Wohlmann P. and Wilson R.		
The WashU-Merck EST Project		
Unpublished (1995)		
Contact: Wilison RK		
Washington University School of Medicine		
4444 Forest Park Parkway, Box B501, St. Louis, MO 63108		
Tel.: 314 286 1800		
Fax: 314 286 1810		
Email: est@watson.wustl.edu		
Insert Size: 766		
High quality sequence stops: 341 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.		
Insert Length: 766 Std Error: 0.00		
Seq primer: -21ml3		
High quality sequence stop: 341.		
Location/Qualifiers		
1..363		
/organism="Homo sapiens"		
/db_xref="GDB:463693"		
/db_xref="taxon:9606"		
/clone="IMAGE:66988"		
/clone_lib="Soares fetal liver spleen INFLS"		
/sex="male"		
/dev_stage="20 week-post conception fetus"		
/lab_host="DH10B (ampicillin resistant)"		
/note="Organ: Liver and Spleen; Vector: pTY73D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5'; AACTCGAAGATTAAATTAAGACTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTY73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."		
BASE COUNT 83 a 83 c 87 g 107 t 3 others		
ORIGIN		
Query Match 65.7%; Score 329; DB 14; Length 363;		
Best Local Similarity 97.3%; Pred. No. 3.66-96;		
Matches 354; Conservative 0; Mismatches 8; Indels 2; Gaps 2;		
17 CCAACAGCCTTAGCTTTTCTGTGGAGATGGGGACCTGAGAGACGACTTGCTGATT 76		
Db 363 CCMAACAGCATTTAGCTTTTCCGTGGAGATGGGGACCTGAGAGACGACTTGCTGATT 304		
Oy 77 TCCTAGGCGTGAATAAGATCCCTAGAGCTAGGTGGGAAGTGACCAACATCCACTTTAAA 136		
Db 303 TCCTAGGCGTGAATAAGATCCCTAGAGCTAGGTGGGAAGTGACCAACATCCACTTTAAA 244		
Oy 137 CACGGGGCTTGCACTTAGCTCACACCTGACCACATCAGAAGCTCCTAAAAATGTAAATT 196		
Db 243 CACGGGGCTTGCACTTAGCTCACACCTGACCACATCAGAAGCTCCTAAAAATGTAAATT 184		
Oy 197 AGGCAGAACAGAGAGCTAAAGAAATGCCAATCATCTATTGCTTAGAGACACAGACGAG 256		

Db 183 AGCAAAAGACAGAGGTAAAGAAATAGCCATATCTATTGCTGAGACAGACAGAG 124

QY 257 GGAACATGATCGGAGATATAAACCAGTCTTCAGCCGCAACGGCAACCCCTTTGGGT 316

Db 123 GGAACAAGATCGGATATTAACCCAGTCTTGA-NCGGCAACGGCAAAACCCCTTTGGG- 66

QY 317 CCCCTCCCTTTGATGAGAGCTCTGTTTTCATGCTATTTCATCTATTAAATCTTGAAC 376

Db 65 CCCCTCCCTTTGATGAGAGCTCTGTTTTCATGCTATTTCATCTATTAAATCTTNAAG 6

QY 377 TGCA 380

Db 5 NACA 2

RESULT 14

AA860368/c 342 bp mRNA linear EST 31-DEC-1998

LOCUS aJ59c05.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394600

DEFINITION 3' similar to contains PRR7.c1 PTR5 repetitive element ;, mRNA sequence.

ACCESSION AA860368.1 GI:2954363

VERSION AA860368.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 342)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapds-remail.nih.gov

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bernaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

Cloning Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www.bio.lnl.gov/dbtrp/image/image.html

Insert Length: 1305 Std Error: 0.00

Seq primer: -40m13 fwd. RT from Amersham

High quality sequence stop: 319.

Location/Qualifiers

1. 342

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1394600"

/clone_1lb="Soares_testis_NHT"

/sex="male"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' GTTACCAATCTGAGAGGAGCGCCGCCAATTCTTTTCTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5 and was constructed by Bento Soares and M. Fatima Bernaldo.

BASE COUNT 75 a 78 c 82 g 107 t

ORIGIN

Query Match 65.2%; Score 326.6; DB 9; Length 342;

Best Local Similarity 98.8%; Pred. No. 2,2e-95;

Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 48 GGGAGTGAAGACAGAGCTAGTGAATTTCTTGAAGTGAAGTCCCTAAGCTTAA 107

Db 342 GGGAGTGAAGACAGAGCTAGTGAATTTCTTGAAGTGAAGTCCCTAAGCTTAA 283

QY 108 CTGGGAAGGTGACACATCCACCTTTTAAACAGCGGGCTTCACTTAGCTACAGCTGAC 167

Db 282 CTGGGAAGGTGACACATCCACCTTTTAAACAGCGGGCTTCACTTAGCTACAGCTGAC 223

QY 168 CAATCAGAGCTCAGCTAAATGCTAAATTGCGAAGACAGAGGTAAAGAAATAGCCAA 227

Db 222 CAATCAGAGCTCAGCTAAATGCTAAATTGCGAAGACAGAGGTAAAGAAATAGCCAA 163

QY 228 TCATCTATTGCTGAGAGCAGAGAGGAGCAATGATCGGAGTATAAACCAGTCTT 287

Db 162 TCATCTATTGCTGAGAGCAGAGAGGAGCAATGATCGGAGTATAAACCAGTCTT 103

QY 288 CGAGCCGCAACGCAACCCCTTTGGGTCCCTCCCTTTGATGAGAGCTGTTTTC 347

Db 102 CGAGCGGCAACGCAACCCCTTTGGGTCCCTCCCTTTGATGAGAGCTGTTTTC 43

QY 348 TGCTATTTCATCTATTAAATCTTTCAGCTGCA 380

Db 42 TGCTATTTCATCTATTAAATCTTTCAGCTGCA 10

RESULT 15

R27412/c 424 bp mRNA linear EST 24-APR-1995

LOCUS yH46d11.s1 Soares_placenta_Nb2HP Homo sapiens cDNA clone

DEFINITION IMAGE:132769 3', mRNA sequence.

ACCESSION R27412.1 GI:783547

VERSION R27412.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 424)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE The WashU-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Insert Size: 695

High quality sequence stops: 367

Source: IMAGE Consortium, LINL

This clone is available royalty-free through LINL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 695 Std Error: 0.00

Seq primer: Promega -21m13

High quality sequence stop: 367.

Location/Qualifiers

1. 424

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:132769"

/clone_1lb="Soares_placenta_Nb2HP"

/sex="Female"

/dev_stage="placenta obtained at birth (full term)"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGGCCGACAGAAATTTTCTTTTCTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization.

BASE COUNT 92 a 96 c 115 g 111 t constructed by Bento Soares and M.Fatima Bonaldo. "
ORIGIN

Query Match 65.1%; Score 326.4; DB 14; Length 424;
Best Local Similarity 93.9%; Pred. No. 2.8e-95;
Matches 351; Conservative 0; Mismatches 21; Indels 2; Gaps 1;

QY 4 TCGGCGAACCTCCGCAAGACAGCCTTAGGTTTCCCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 373 TCGATCAACCTCCGCAAGACAGCCTTAGGTTTCCCTGTTGAGATGGGGAGCTGAGAGACAG 314
QY 64 GACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGAAAGTGACAC 123
DB 313 GACTAGCTGATTTCTTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGAAAGTGACTGC 254
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAC 183
DB 253 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAGACCTGACCAATC--AGAGCTCAC 196
QY 184 TAAATGCTAATTAAGCAAGAAGAGAGTAAAGAATAGCCAAATCATCTATTGCTGAG 243
DB 195 TAAATGCTAATTAAGCAAGAAGAGAGTAAAGAATAGCCAAATCATCTATTGCTGAG 136
QY 244 AGCAGCAGCAGAGGGGCAATGATCGGATATATAACCAAGTCTTGAGCCGGCAACGGCA 303
DB 135 AGCAGCAGCAGAGGGGCAAGATCGGATATATAACCAAGTCTTGAGCCGGCAACGGCA 76
QY 304 ACCGCTTTGGGTCCCTCCCTTTGTATGGAGCTTGTGTTCAATGCTATTTCACCTAT 363
DB 75 AACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTTGTGTTCAATGCTATTTCACCTAT 16
QY 364 TAAATCTGCAACT 377
DB 15 TAAATCTGCAACT 2

Search completed: April 17, 2003, 07:35:33
Job time : 1270.84 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 10:07:23 ; Search time 1703.47 Seconds

(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Sequence: 1 caagatccgcagatcataca.....tcagtgcagacacatccag 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```
GenBank!:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_ph:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	10499	6	AX007980
2	501	100.0	56093	6	AX329572
3	501	100.0	56093	9	HSAC000064
4	501	100.0	149194	9	AC007566
5	474	94.6	2938	6	AX000957
6	474	94.6	2938	6	AX027471
7	474	94.6	2938	6	AF072499
8	442	88.2	7582	6	AX000966
9	442	88.2	7582	6	AX027480
10	435	86.8	40205	9	AF045450
11	435	86.8	142742	9	AF121782
12	435	86.8	340000	9	HS21C080
13	428.6	85.5	8339	6	AL162912
14	428.4	85.5	3372	6	AX000964
15	428.4	85.5	3372	6	AX027478
16	428.4	85.5	3372	6	AF072502
17	425.4	84.9	146545	2	AC104009
18	423.8	84.6	143063	9	AC040936
19	423.8	84.6	163166	9	AL356632
20	423.8	84.6	176773	2	AC009867
21	422.2	84.3	2304	14	AF009668
22	422.2	84.3	108232	9	AP000654
23	422.2	84.3	156938	9	AL159163
24	422.2	84.3	184523	2	AP002890
25	421.2	84.1	139744	9	AL133513
26	420.6	84.0	132981	9	HS82J11
27	420.6	84.0	161771	9	CNS01DRX
28	420.6	84.0	178152	2	AC108746
29	420.6	84.0	191426	2	AC026086
30	420.6	84.0	194545	2	AC087482
31	420.6	84.0	194718	2	AC026977
32	420.6	84.0	220807	9	AC093334
33	420.6	84.0	251124	9	HDA5000660
34	419.6	83.8	1158	6	A46012
35	419.6	83.8	1158	6	A57048
36	419.6	83.8	1158	6	A60113
37	419.6	83.8	1158	6	A79463
38	419.6	83.8	1158	6	A80197
39	419.6	83.8	1158	6	A80306
40	419.6	83.8	1158	6	AR035349
41	419.6	83.8	1158	6	AR036190
42	419.6	83.8	1158	6	AR036710
43	419.6	83.8	1158	6	AR076972
44	419.6	83.8	1158	6	AR094693
45	419.6	83.8	1158	6	AR183928

ALIGNMENTS

RESULT 1
AX007980
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AX007980.1 GI:9995677

Sequence 3 from Patent WO9967395. 10499 bp. DNA linear PAT 06-SEP-2000

REFERENCE
AUTHORS
TITLE
1 (bases 1 to 10499)
Perin,J.P., Rieger,F. and Alliel,P.M.
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses

JOURNAL Patent: WO 9667395-A 3 29-DEC-1999;
INST NAT SANTE RECH MED (FR); FERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES Location/Qualifiers
SOURCE 1..10499
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 2195 t
ORIGIN

Query Match 100.0%; Score 501; DB 6; Length 10499;
Best Local Similarity 100.0%; Pred. No. 2.3e-138;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAGATCTCAGGATTTATCATGAGGCTGTTGTTCTCTATACCGACGTGTACCTTACGCT 60
Db 5000 CAGATCTCAGGATTTATCATGAGGCTGTTGTTCTCTATACCGACGTGTACCTTACGCT 5059
Qy 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 5060 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
Qy 121 GATGCTCTCTGTGATCCCTGATCACTGCTGCTCAATTTCTTTGTTGCTTTGAAGAT 180
Db 5120 GATGCTCTCTGTGATCCCTGATCACTGCTGCTCAATTTCTTTGTTGCTTTGAAGAT 5179
Qy 181 ACTCAAAACCAACATCTCACTGACCTGACCTATTTTACCCCAAGGGTTGAGGATAGT 240
Db 5180 ACTCAAAACCAACATCTCACTGACCTGACCTATTTTACCCCAAGGGTTGAGGATAGT 5239
Qy 241 CCCCATCTAATTTGGCCAGGCAATTAAGCCCAAGACTTGAGCAATCTCTATACCTTGACACT 300
Db 5240 CCCCATCTAATTTGGCCAGGCAATTAAGCCCAAGACTTGAGCAATCTCTATACCTTGACACT 5299
Qy 301 TGTCCTTGAGTGTGATGATTTACTTTTGCCCTCCCAATTAAGAAACCTTTGCTATCA 360
Db 5300 TGTCCTTGAGTGTGATGATTTACTTTTGCCCTCCCAATTAAGAAACCTTTGCTATCA 5359
Qy 361 AGCCACCCCAAGCGCTCTTCAATTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 5360 AGCCACCCCAAGCGCTCTTCAATTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5419
Qy 421 GGCTCAACTCTGCTGACAGAGGTTACTTAAAGGCTTAAATTTATCCAAAGGACCAAGGGCC 480
Db 5420 GGCTCAACTCTGCTGACAGAGGTTACTTAAAGGCTTAAATTTATCCAAAGGACCAAGGGCC 5479
Qy 481 CTCAGTGAGGAACATCTCAG 501
Db 5480 CTCAGTGAGGAACATCTCAG 5500

RESULT 2
AX329572 56093 bp DNA linear PAT 09-JAN-2002
LOCUS AX329572
DEFINITION Sequence 81 from Patent WO0194629.
ACCESSION AX329572
VERSION AX329572.1 GI:18102550
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Hortigian, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;
Avalon Pharmaceuticals (US)

FEATURES
SOURCE location/Qualifiers
1..56093
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match 100.0%; Score 501; DB 6; Length 56093;
Best Local Similarity 100.0%; Pred. No. 2.3e-138;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAGATCTCAGGATTTATCATGAGGCTGTTGTTCTCTATACCGACGTGTACCTTACGCT 60
Db 33000 CAGATCTCAGGATTTATCATGAGGCTGTTGTTCTCTATACCGACGTGTACCTTACGCT 33059
Qy 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 33060 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 33119
Qy 121 GATGCTCTCTGTGATCCCTGATCACTGCTGCTCAATTTCTTTGTTGCTTTGAAGAT 180
Db 33120 GATGCTCTCTGTGATCCCTGATCACTGCTGCTCAATTTCTTTGTTGCTTTGAAGAT 33179
Qy 181 ACTCAAAACCAACATCTCACTGACCTGACCTATTTTACCCCAAGGGTTGAGGATAGT 240
Db 33180 ACTCAAAACCAACATCTCACTGACCTGACCTATTTTACCCCAAGGGTTGAGGATAGT 33239
Qy 241 CCCCATCTAATTTGGCCAGGCAATTAAGCCCAAGACTTGAGCAATCTCTATACCTTGACACT 300
Db 33240 CCCCATCTAATTTGGCCAGGCAATTAAGCCCAAGACTTGAGCAATCTCTATACCTTGACACT 33299
Qy 301 TGTCCTTGAGTGTGATGATTTACTTTTGCCCTCCCAATTAAGAAACCTTTGCTATCA 360
Db 33300 TGTCCTTGAGTGTGATGATTTACTTTTGCCCTCCCAATTAAGAAACCTTTGCTATCA 33359
Qy 361 AGCCACCCCAAGCGCTCTTCAATTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 33360 AGCCACCCCAAGCGCTCTTCAATTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33419
Qy 421 GGCTCAACTCTGCTGACAGAGGTTACTTAAAGGCTTAAATTTATCCAAAGGACCAAGGGCC 480
Db 33420 GGCTCAACTCTGCTGACAGAGGTTACTTAAAGGCTTAAATTTATCCAAAGGACCAAGGGCC 33479
Qy 481 CTCAGTGAGGAACATCTCAG 501
Db 33480 CTCAGTGAGGAACATCTCAG 33500

RESULT 3
HSAC000064 56093 bp DNA linear PRI 13-NOV-1996
LOCUS HSAC000064
DEFINITION Human BAC clone RG083M05 from 7q21-7q22, complete sequence.
ACCESSION AC000064
VERSION AC000064.1 GI:1669369
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 56093)
AUTHORS Paulley, A.
TITLE The sequence of H. sapiens BAC clone RG083M05
JOURNAL Unpublished (1996)
REFERENCE 2 (bases 1 to 56093)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1996)
COMMENT Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
e-mail: saplens@wustl.edu

NOTICE: This sequence may not represent the entire insert of this
sections once, or longer because we only sequence overlapping
neighborings submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The
library contains cloned DNA from a human male fibroblast cell line
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBELO

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this
clone is at base position 1 of H_RG083M05; actual end is at 56093
of H_RG083M05

This clone contains STS SMS1725.

Location/Qualifiers

1. 56093

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone="H_RG083M05"

/clone_id="CITB-978SK-B"

/complement(838..1131)

repeat_region

gene

/gene="WUGSC:H_RG083M05.1"

/rpt_family="ALU"

CDS

join(<1360..1503,4181..4370,4587..4774,6422..6556,
9483..9547,11631..11773,11864..12021,13131..13296,
14885..14988,16349..16546,16837..16971)
/gene="WUGSC:H_RG083M05.1"

/note="Atpase; strong similarity to peroxisome
biosynthesis protein PAB1 (PID:g1172019); coded for by
human cDNA C04279 (NID:g1467550)"

/codon_start=1

/protein_id="PAB46346.1"

/db_xref="GI:1669371"

/translation="KRLEINQKTELEVAFAVMNOPSVLDDLDLIAGLPAVEHEH
SPABORCEILNVIKKIDCDINKFTDLDLGVAKEGTGFVARDPVLVDRATHRL
SRGSISTRKLVTTTIDPQALRGFPAPASIRSNLHKRPLGMDKIGGHEHQIILMD
TIDLPKAVCKKKEYPBLFANLPFIRQRTGTLGPPETGKTLGAVIARSRNPFISV
KGBELSKYIGASEQAVRDIFRAQAKPCILFFDEESIAPRGHDTGTVRVNQ
LTLQDVEGLQGVVLAATSRDLIDPALLRGRLDCVCPDPQVITSYLESKTQ
OMHSFVSRLEILINVLSGLPLADVDLQHVASVYDSFGADLKALVLAOLEALHG
MLSKSEILPDESKFMVRLYFGSSYVESLNGTSDLSGLSA PSSAWTODLPVP
GKQDLSQPRVLTASQEGCELTQBRDLPRLDISIKRRTYSGSGEDSMNQPPI
KIRLAISQSHLMTALGHTRPISIEDDKNFAEL"

repeat_region

/complement(4948..5130)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(6581..7133)

repeat_region

/rpt_family="U1"

repeat_region

/complement(7767..8037)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(8186..8472)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(8625..8738)

repeat_region

/note="match to human 3' EST H75782 (NID:g1049794), bases
287-444"

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 5' EST H75921 (NID:g1050050), bases
21-348"

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 5' EST N2627 (NID:g1130501), bases
276-343"

repeat_region

/complement(12612..12907)

repeat_region

/rpt_family="ALU"

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 5' EST H41382 (NID:g17434), bases
143-266"

repeat_region

/rpt_family="ALU"

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 5' EST H41382 (NID:g17434), bases
30-58"

repeat_region

/rpt_family="ALU"

repeat_region

/complement(14110..14137)

repeat_region

/rpt_family="U1"

repeat_region

/complement(15618..15907)

repeat_region

/rpt_family="ALU"

repeat_region

/rpt_family="ALU"

repeat_region

/note="match to human fetal brain 5' EST D61494
(NID:g970409), bases 1-255, and to human 3' EST R07476
(NID:g975399)"

repeat_region

/rpt_family="ALU"

repeat_region

/note="similarity to various SS-RNA virus polyproteins;
pseudogene; region of matches and close matches to
multiple human ESTs, see R68740 (NID:g942257)"

repeat_region

/rpt_family="ALU"

repeat_region

/note="Grail prediction, score = 80"

repeat_region

/evidence=not experimental

repeat_region

/complement(38938..39224)

repeat_region

/rpt_family="ALU"

repeat_region

/note="match to multiple human ESTs, see N30113
(NID:g1148633)"

repeat_region

/rpt_family="ALU"

repeat_region

/complement(40247..40538)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(40632..40924)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(42283..42891)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(45474..45613)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(45614..45737)

repeat_region

/note="match to human 3' EST H48898 (NID:g988738), bases
129-333"

repeat_region

/complement(46107..47026)

repeat_region

/note="match to multiple human ESTs, see N81064
(NID:g1243765), H48897 (NID:g988737), and M78831
(NID:g273146)"

repeat_region

/complement(47027..47318)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(47365..47782)

repeat_region

/note="match to multiple human ESTs, see W37495
(NID:g1319089)"

repeat_region

/rpt_family="ALU"

repeat_region

/complement(48116..48405)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(48406..48584)

repeat_region

/note="match to human 3' EST N29952 (NID:g1148472), bases
230-455, and 5' EST R12730 (NID:g765806)"

repeat_region

/complement(48787..49405)

repeat_region

/rpt_family="ALU"

repeat_region

/complement(49406..49534)

repeat_region

/note="match to human 3' EST R65794 (NID:g838432), bases
309-440"

```

repeat_region complement(49638, .49672)
             /rpt family="ALU"
misc_feature /note="match to human 3' EST N29952 (NTD:g1148472) and 5'
             EST N2938 (NTD:g1148158), sequences are from opposite
             ends of the same clone"
gene         /gene="WUGSC:H.RG083M05.2"
             /complement(49698, .51306)
             /note="WUGSC:H.RG083M05.2"
             /complement(join(49698, .49888,51575, .51806))
             /gene="WUGSC:H.RG083M05.2"
             /note="coded for by human cDNA M7389 (NTD:g119205),
             R65891 (NTD:g838529), R65794 (NTD:g838432) and R65794
             (NTD:g838432)"
CDS          /codon_start=1
             /protein_id="AAB46345.1"
             /db_xref="GI:1569370"
             /translation="MFYRFGCGIIFPCPGVYVQIGNVSVIDEQGRPYAIRGF
             IDQYCEKSAALTWLITPLSSPRQPDASVITGEEDLPKMBLEFVCHAPSEYRK
             SRSPFPVTPRPEKGYIWHV3PTPALTIKESVANHL"
exon         /complement(51576, .51738)
             /gene="WUGSC:H.RG083M05.2"
             /note="Grail prediction, score = 86"
repeat_region /evidence=not_experimental
             /complement(52052, .52329)
             /rpt family="L1"
             /note="match to human HST M79192 (NTD:g273505) base 2-289"
misc_feature 55557, .55843
             /note="match to human HST M79192 (NTD:g273505) base 2-289"

Query Match 100.0%; Score 501, DB 9, Length 56093;
Best Local Similarity 100.0%; Pred. No. 2,3e-138;
Matches 501, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

Qy 1 CAAGTCTCAGATTATCATGAGCTGTGTCCTCATATACCAAGCTGTACTGACCT 60
Db 33000 CAGATCTCAGATTATCATGAGCTGTGTCCTCATATACCAAGCTGTACTGACCT 33059
Qy 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGTGTTTACAGTCTCTGACCTTCAG 120
Db 33060 TATACCTGCTTTCCCAATATACAGAGAGAGAGTGTTTACAGTCTCTGACCTTCAG 33119
Qy 121 GATGCTCTTCTTGCACTCCCTGTACATCTGACTTCAATTTCTTTGGCTTTGAAGT 180
Db 33120 GATGCTCTTCTTGCACTCCCTGTACATCTGACTTCAATTTCTTTGGCTTTGAAGT 33179
Qy 181 ACTTAAACCCCAACATCTCACTGAGACTATTTTACCCCAAGGTTTCCAGGATAGT 240
Db 33180 ACTTAAACCCCAACATCTCACTGAGACTATTTTACCCCAAGGTTTCCAGGATAGT 33239
Qy 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGACTTACGCAATCTCTCATACCTGACACT 300
Db 33240 CCCCATCTATTTGGCCAGGATTAAGCCCAAGACTTACGCAATCTCTCATACCTGACACT 33299
Qy 301 TGTCTTGGTGAAGTGAATGATTTACTTTGGCCGTCATTAGAAACCTTGTGCCATCA 360
Db 33300 TGTCTTGGTGAAGTGAATGATTTACTTTGGCCGTCATTAGAAACCTTGTGCCATCA 33359
Qy 361 AGCCACCCCAAGGCTCTTCAATTTCTGCTACTCTGTGCTACATGAGTTTCCAAACAA 420
Db 33360 AGCCACCCCAAGGCTCTTCAATTTCTGCTACTCTGTGCTACATGAGTTTCCAAACAA 33419
Qy 421 GGCTCAACTGTGCTCAAGCAGGTTACTTGAAGCTAAATATCCAAAGCAGCAGGGCC 480
Db 33420 GGCTCAACTGTGCTCAAGCAGGTTACTTGAAGCTAAATATCCAAAGCAGCAGGGCC 33479
Qy 481 CTCAGTGAAGAACATCTCAG 501
Db 33480 CTCAGTGAAGAACATCTCAG 33500

```

```

RESULT 4
AC007566/c AC007566 149194 bp DNA linear PRI 01-MAR-2002
LOCUS Homo sapiens BAC clone CTB-1005 frcm 7q21-7q22, complete sequence.
DEFINITION

```

```

ACCESSION AC007566
VERSION AC007566.2 GI:11181861
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
          1 (bases 1 to 149194)
          Sultion,J.E. and Waterston,R.
          Toward a complete human genome sequence
          Genome Res. 8 (11), 1097-1108 (1998)
          99063792
          PUBMED 9847074
REFERENCE 2 (bases 1 to 149194)
          Du,Z.
          The sequence of Homo sapiens BAC clone CTB-1005
          Unpublished (2001)
          3 (bases 1 to 149194)
REFERENCE Waterston,R.H.
          Direct Submission
          Submitted (15-MAY-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          4 (bases 1 to 149194)
REFERENCE Waterston,R.
          Direct Submission
          Submitted (02-OCT-2000) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          5 (bases 1 to 149194)
REFERENCE Waterston,R.H.
          Direct Submission
          Submitted (16-NOV-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          6 (bases 1 to 149194)
REFERENCE Waterston,R.H.
          Direct Submission
          Submitted (03-JAN-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          7 (bases 1 to 149194)
REFERENCE Waterston,R.H.
          Direct Submission
          Submitted (06-FEB-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          8 (bases 1 to 149194)
REFERENCE Waterston,R.
          Direct Submission
          Submitted (01-MAR-2002) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          On Nov 16, 2000 this sequence version replaced gi:4835815.
          ----- Genome Center
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc
          Contact: sapiens@wustl.wustl.edu
          ----- Summary Statistics
          Center project name: H_RG010G05

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

source

```
1..149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_1lb="CTB-978SK-B"
1..1634
/rpt_family="L1"
misc_feature
2248..2388
/note="match to EST BG752883 (NID:g14063536)"
misc_feature
2248..2387
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
2248..2387
/note="match to EST AW579261 (NID:g7254310)"
misc_feature
2248..2387
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
2248..2387
/note="match to EST B1160365 (NID:g14620366)"
misc_feature
2248..2374
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
2248..2287
/note="match to EST AV686676 (NID:g10288539)"
misc_feature
2253..2387
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
2344..2387
/note="match to EST AV686676 (NID:g10288539)"
repeat_region
2696..3066
/rpt_family="MAIR"
repeat_region
3108..3392
/rpt_family="Alu"
3540..3628
/note="match to EST AA425526 (NID:g12106267) zw48b03.r1"
misc_feature
3540..3628
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
3542..3632
/note="match to EST AV686676 (NID:g10288539)"
misc_feature
3542..3628
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
3542..3628
/note="match to EST AU123510 (NID:g10948226)"
3542..3628
/note="match to EST BG766882 (NID:g14077535)"
```

```
misc_feature 3542..3628
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 3542..3609
/note="match to EST AW579261 (NID:g7254310)"
misc_feature 3562..3628
/note="match to EST BE27264 (NID:g9146913)"
misc_feature 3598..3628
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3714..3785
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 3717..3785
/note="similar to Mus musculus EST BE994936 (NID:g10678674)"
misc_feature 3719..3785
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 3719..3785
/note="match to EST AA425526 (NID:g12106267) zw48b03.r1"
misc_feature 3719..3785
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 3719..3785
/note="match to EST BE27264 (NID:g9146913)"
misc_feature 3719..3785
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3719..3785
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 3719..3785
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 3719..3785
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature 3860..3862
/note="match to EST AV686676 (NID:g10288539)"
repeat_region 4730..4827
/rpt_family="L2"
4834..4925
/rpt_family="L1"
misc_feature 5715..5860
/note="match to EST AA425526 (NID:g12106267) zw48b03.r1"
misc_feature 5715..5860
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 5715..5860
/note="match to EST BE27264 (NID:g9146913)"
misc_feature 5715..5860
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 5715..5860
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 5715..5860
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 5715..5860
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 5715..5860
/note="similar to Mus musculus EST BE994936

Query Match 100.0%; Score 501; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 2.2e-138;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGATTCAGATATCATATGAGGCTGTTCTTATAGCCAGCTGATACGCTT 60
DB 89004 CAAGATTCAGATATCATATGAGGCTGTTCTTATAGCCAGCTGATACGCTT 88945
QY 61 TATATCTGCTTCCCAATATACAGAGAGAGAGGTTTACAGTCCGACCTTCA 120
DB 88944 TATATCTGCTTCCCAATATACAGAGAGAGAGGTTTACAGTCCGACCTTCA 88985
QY 121 GATGCTTCTTTCGATCCCTGATACATCTGACTCTCAATTTCTTGTGCTTGAAGAT 180
DB 88884 GATGCTTCTTTCGATCCCTGATACATCTGACTCTCAATTTCTTGTGCTTGAAGAT 88825
QY 181 ACTTCAACCAACATTCATCACTGACATATTTTACCCCAAGGTTCAAGGATAGT 240
DB 88824 ACTTCAACCAACATTCATCACTGACATATTTTACCCCAAGGTTCAAGGATAGT 88765
```

241 |CCCCATCTATTGGGCGAGGACATTAGCCCAAGAC'TGAGCCCAATCCTCATACCTGGAGACT| 300
|CCCCATCTATTGGGCGAGGACATTAGCCCAAGAC'TGAGCCCAATCCTCATACCTGGAGACT| 88705
301 |TGTCTCTGGTAGGTGATGATTTTATCTTTGGCGGCCCATTCAGAAACCTTGTGCTATCA| 360
88704 |TGTCTCTGGTAGGTGATGATTTTATCTTTGGCGGCCCATTCAGAAACCTTGTGCTATCA| 88645
361 |AGCCACCCCAAGGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAACCAAA| 420
88644 |AGCCACCCCAAGGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAACCAAA| 88585
421 |GGCTCAACTGTCTGACAGAGGTTACTTAGGGCTAAATTTCCAAAGGACACAGGGCC| 480
88584 |GGCTCAACTGTCTGACAGAGGTTACTTAGGGCTAAATTTCCAAAGGACACAGGGCC| 88525
481 |CTCAGTGAGGAACACATCCAG| 501
88524 |CTCAGTGAGGAACACATCCAG| 88504

RESULT 5
AX000957 2938 bp DNA linear PAT 10-MAR-2000
LOCUS Sequence 2 from Patent WO9902696
DEFINITION AX000957
ACCESSION AX000957
VERSION AX000957.1 GI:7241199
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
unclassified.
1 (bases 1 to 2938)
REFERENCE
AUTHORS Beseme, F. and Blond, J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 9902696-A 2 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source location/Qualifiers
1..2938
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN

Query Match 94.6%; Score 474; DB 6; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

1 CAAGATCTCAGGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGCTGACTAGCCC 59
2079 CAAAGATCTCAGGATTATCAATGAGGCGCTTGTCTTTATACCAAGCTGATACAGCCC 2138
60 TTATACCTGCTTTTCCCAATATACAGAGAGACAG;GTGGTTTACAGTCTGAGCCTTCA 119
2139 TTATACCTGCTTTTCCCAATATACAGAGAGACAG;GTGGTTTACAGTCTGAGCCTTCA 2198
120 GGAATGCTTTCTTGTGATCCCTGTATACCTGATCTTCAATTTCTTTGGTCCCTTTGAAGA 179
2199 GGAATGCTTTCTTGTGATCCCTGTATACCTGATCTTCAATTTCTTTGGTCCCTTTGAAGA 2258
180 TACTTCAAAACCAACATCTCACTACCTGACTATTTTACCCCAAGGTTCAAGGATAG 239
2259 TACTTCAAAACCAACATCTCACTACCTGACTATTTTACCCCAAGGTTCAAGGATAG 2318
240 TCCCATCTATTGGCGAGGATTAAGCCCAAGACTT3AGCCCAATCTCTCATACCTGAGAC 299
2319 CCCCATCTATTGGCGAGGATTAAGCCCAAGACTT3AGCCCAATCTCTCATACCTGAGAC 2378
300 TTGTCTCTTGGTAGGTGATGATTTTATCTTTGGCGGCCCATTCAGAAACCTTGTGCTATC 359
2379 TTGTCTCTTGGTAGGTGATGATTTTATCTTTGGCGGCCCATTCAGAAACCTTGTGCTATC 2438
360 AAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAACCAA 419

2439 |AAGCACCACCAAGCGCTCTTCAATTTCTCTGCTACCTGTGGCTACATGTTTCCAAACCAA| 2498
420 |AGGCTCAACTCTGCTACAGAGGTTACTTAGGGCTAAATTTATCCAAAGACACAGGGC| 479
2499 |AGGCTCAACTCTGCTACAGAGGTTACTTAGGGCTAAATTTATCCAAAGACACAGGGC| 2558
480 |CCTCAGTAGGAACACATCCAG| 501
2559 |CCTCAGTAGGAACACATCCAG| 2580

RESULT 6
AX027471 2938 bp DNA linear PAT 16-SEP-2000
LOCUS AX027471
DEFINITION Sequence 21 from Patent FR2788784.
ACCESSION AX027471
VERSION AX027471.1 GI:10188435
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Maillet, F., Voisinet, C. and Paranhos, B.G.
JOURNAL Patent: FR 2788784-A 21 28-JUL-2000;
BIO MERIEUX (FR)
FEATURES
source location/Qualifiers
1..2938
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN

Query Match 94.6%; Score 474; DB 6; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

1 CAAGATCTCAGGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGCTGACTAGCCC 59
2079 CAAAGATCTCAGGATTATCAATGAGGCGCTTGTCTTTATACCAAGCTGATACAGCCC 2138
60 TTATACCTGCTTTTCCCAATATACAGAGAGACAGAGTGTTTACAGTCTGAGCCTTCA 119
2139 TTATACCTGCTTTTCCCAATATACAGAGAGACAGAGTGTTTACAGTCTGAGCCTTCA 2198
120 GGAATGCTTTCTTGTGATCCCTGTATACCTGATCTTCAATTTCTTTGGTCCCTTTGAAGA 179
2199 GGAATGCTTTCTTGTGATCCCTGTATACCTGATCTTCAATTTCTTTGGTCCCTTTGAAGA 2258
180 TACTTCAAAACCAACATCTCACTACCTGACTATTTTACCCCAAGGTTCAAGGATAG 239
2259 TACTTCAAAACCAACATCTCACTACCTGACTATTTTACCCCAAGGTTCAAGGATAG 2318
240 TCCCATCTATTGGCGAGGATTAAGCCCAAGACTTGAAGCAATCTCTCATACCTGAGAC 299
2319 CCCCATCTATTGGCGAGGATTAAGCCCAAGACTTGAAGCAATCTCTCATACCTGAGAC 2378
300 TTGTCTCTTGGTAGGTGATGATTTTATCTTTGGCGGCCCATTCAGAAACCTTGTGCTATC 359
2379 TTGTCTCTTGGTAGGTGATGATTTTATCTTTGGCGGCCCATTCAGAAACCTTGTGCTATC 2438
360 AAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAACCAA 419
2439 AAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAACCAA 2498
420 AAGCTCAACTCTGCTACAGAGGTTACTTAGGGCTAAATTTATCCAAAGACACAGGGC 479
2499 AAGCTCAACTCTGCTACAGAGGTTACTTAGGGCTAAATTTATCCAAAGACACAGGGC 2558
480 CCTCAGTAGGAACACATCCAG 501
2559 CCTCAGTAGGAACACATCCAG 2580

RESULT 7
AF072499 2938 bp mRNA linear PRI 10-FEB-1999
LOCUS Homo sapiens endogenous retrovirus W sequence.
DEFINITION AF072499
ACCESSION AF072499.1 GI:4262283
VERSION
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 2938)
AUTHORS Blond, J. L., Beseme, F., Duret, L., Bouton, O., Bedin, F., Perron, H., Mandrand, B. and Mallet, F.
TITLE Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family
JOURNAL J. Virol. 73 (2), 1175-1185 (1999)
MEDLINE 99099005
PUBMED 9882319
REFERENCE 2 (bases 1 to 2938)
AUTHORS Blond, J. L., Beseme, F. and Mallet, F.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1998) UM103 CNRS-biomerieux, ENS Lyon, 46 allée d'Italie, Lyon, Cedex 07 69364, France
FEATURES
source location/Qualifiers
1..2938 /organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Cl.6A1"
/cissue_type="placenta"
/clone_lib="Clontech 5'-stretch plus library, Cat number HL0143"
misc_feature 1..2938 /note="corresponds to gag and pol"
repeat_region 1..2938 /rpt_family="HERV-W"
BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN
Query Match 94.6%; Score 474; DB 9; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 1 CAAGATCTCAGGATTTAAT-GAGGCTGTGTCTCTATAGCCAGTGAAGCC 59
DB 2079 CAAGATCTCAGGATTTAATGAGGCGGTGTCTTTATACCACTGAAGCC 2138
QY 60 TTATATCTGTGCTTCCCAATACAGAGAGAGAGGTTTACAGTCTGAAGCTTCA 119
DB 2139 TTATATCTGTGCTTCCCAATACAGAGAGAGAGGTTTACAGTCTGAAGCTTCA 2138
QY 120 GGATGCTTCTTCTGATCCCTGTACATCTGACTCTCAATCTTGTGTGCTTGAAGA 179
DB 2199 GGATGCTTCTTCTGATCCCTGTACATCTGACTCTCAATCTTGTGTGCTTGAAGA 2258
QY 180 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGGTTCAGGGATAG 239
DB 2259 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGGTTCAGGGATAG 2318
QY 240 TCCCATCTATTTGGCCAGGATAGCCCAAGACTTGAGCAATCTCATACCTGACAC 299
DB 2319 CCCCACATATTTGGCCAGGATAGCCCAAGACTTGAGCAATCTCATACCTGACAC 2378
QY 300 TTGTCTCTCGGTAGGTGATATTTTGTGGCCGCCATTCAGAAAAGCTTGTGCAATC 359
DB 2379 TTGTCTCTCGGTAGGTGATATTTTGTGGCCGCCATTCAGAAAAGCTTGTGCAATC 2438
QY 360 AAGGACCCCAAGCGCTTTCATATTTCTGCTACCTGTGGCTACATGTTTCCAAACCA 419
DB 2439 AAGGACCCCAAGCGCTTTCATATTTCTGCTACCTGTGGCTACATGTTTCCAAACCA 2498

QY 420 AGGCTCACTGCTCAGACAGGTTACTTAGGGCTAAATTTATCCAAAGGACAGGGC 479
DB 2499 AGGCTCACTGCTCAGACAGGTTACTTAGGGCTAAATTTATCCAAAGGACAGGGC 2558
QY 480 CCTCAGTGAGGAAACATCCAG 501
DB 2559 CCTCAGTGAGGAAACATCCAG 2580
RESULT 8
AX000966
LOCUS AX000966 7582 bp DNA linear PAT 10-MAR-2000
DEFINITION Sequence 11 from Patent WO9902696.
ACCESSION AX000966
VERSION AX000966.1 GI:7241208
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 7582)
AUTHORS Beseme, F. and Blond, J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 9902696-A 11 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source location/Qualifiers
1..7582 /organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2156 a 1876 c 1538 g 1796 t 216 others
ORIGIN
Query Match 88.2%; Score 442; DB 6; Length 7582;
Best Local Similarity 91.3%; Pred. No. 1e-120;
Matches 464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;
QY 1 CAAGATCTCAGGATTTAAT-GAGGCTGTGTCTCTATAGCCAGTGAAGCC 59
DB 2678 CAAGATCTCAGGATTTAATGAGGCGGTGTCTTTATACCACTGAAGCC 2737
QY 60 TTATATCTGTGCTTCCCAATACAGAGAGAGAGGTTTACAGTCTGAAGCTTCA 119
DB 2738 TTATATCTGTGCTTCCCAATACAGAGAGAGAGGTTTACAGTCTGAAGCTTCA 2797
QY 120 GGATGCTTCTTCTGATCCCTGTACATCTGACTCTCAATCTTGTGTGCTTGAAGA 179
DB 2798 GGATGCTTCTTCTGATCCCTGTACATCTGACTCTCAATCTTGTGTGCTTGAAGA 2857
QY 180 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGGTTCAGGGATAG 239
DB 2858 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGGTTCAGGGATAG 2917
QY 240 TCCCATCTATTTGGCCAGGATAGCCCAAGACTTGAGCAATCTCATACCTGACA- 298
DB 2918 YCCCATCTATTTGGCCAGGATAGCCCAAGACTTGAGCAATCTCATACCTGACA- 2977
QY 299 -CTTGTCTCTCGGTAGGTGATATTTTGTGGCCGCCATTCAGAAAAGCTTGTGCA 357
DB 2978 TCTGTCTCTCGGTAGGTGATATTTTGTGGCCGCCATTCAGAAAAGCTTGTGCA 3037
QY 358 TCAAGCACCAAGCGCTTTCATATTTCTGCTACCTGTGGCTACATGTTTCCAAAC 417
DB 3038 TCAAGCACCAAGCGCTTTCATATTTCTGCTACCTGTGGCTACATGTTTCCAAAC 3057
QY 418 AAGGCTCAACTGTGCTCAGCAGGT---TACTTAGGGCTAAATTTATCCAAAGGAC 473
DB 3098 AARAGCTCACTGTGCTCAGCAGGTTAAATCTTAGGCTAARATTTATCCAAAGGAC 3157
QY 474 CAGGCGCTCAGTGAGGAAACATCCAG 501
DB 3158 CAGGCGCTCAGTGAGGAAATATCCAG 3185

LOCUS	AX027480	7582 bp	DNA	linear	PAT 16-SEP-2000
DEFINITION	Sequence 30 from Patent FR2788784.				
ACCESSION	AX027480				
VERSION	AX027480.1	GI:10188444			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Creniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 7582)				
AUTHORS	Mallet, F., Voisset, C. and Paranhos, B.G.				
JOURNAL	Patent: FR 2788784-A 30 28-JUL-2000;				
	BIO MERIEUX (FR)				
FEATURES	Location/Qualifiers				
source	1..7582				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	2156 a 1876 c 1538 g 1796 t	216 others			
ORIGIN					
Query Match	88.2%; Score 442; DB 6; Length 7582;				
Best Local Similarity	91.3%; Pred. No. 1e-120;				
Matches	464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;				
QY	1 CAAGATCTCAGGATTATCAAT-GAGGCTGTGTCTCTTANAGCAAGCTGTACCTAGCCC	59			
Db	2678 CAAGATCTCAGGATTATCAATGAGGCGGTGTCTCTTATACCAGGCTGTACCTAGCCC	2737			
QY	60 TTATATCTGCTTTTCCCAATACACAGAGAACAGAGTGTGTTTACAGTCTTGACCTTCA	119			
Db	2738 TTATATCTGCTTTTCCCAATACACAGAGAACAGAGTGTGTTTACAGTCTTGACCTTCA	2797			
QY	120 GGATGCTCTCTCTGATCCCTGTACATCTGACTTCAATCTTGTGTGCTTTGAAGA	179			
Db	2798 GGATGCTCTCTCTGATCCCTGTACATCTGACTTCAATCTTGTGTGCTTTGAAGA	2857			
QY	180 TACTTCAACCCCAACATCTCACTGACCTAATTTTACCACCAAGGTTGAGGATAG	239			
Db	2858 TACTTCAACCCCAACATCTCACTGACCTAATTTTACCACCAAGGTTGAGGATAG	2917			
QY	240 TCCCATCTAATTTGGCAGGCAATGACCCCAAGACTGACCAATCTCTTACTTGACA-	298			
Db	2918 TCCCATCTAATTTGGCAGGCAATGACCCCAAGACTGACCAATCTCTTACTTGACA-	2977			
QY	299 -CTTGCTCTCGGTAGTGTGATGATTACTTTTGGCGGCCCATTCAGAAACCTTGCCA	357			
Db	2978 TCTTGCTCTCGGTAGTGTGATGATTACTTTTGGCGGCCCATTCAGAAACCTTGCCA	3037			
QY	358 TCAAGCACCCCAAGCGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAAC	417			
Db	3038 TCAAGCACCCCAAGCGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAAC	3097			
QY	418 AAAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTTAAATTTATCCAAAGGCAC	473			
Db	3098 AAAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTTAAATTTATCCAAAGGCAC	3157			
QY	474 CAGGGCCCTCAGTGGAGAACACATCCAG	501			
Db	3158 CAGGGCCCTCAGTGGAGAACATCCAG	3185			
RESULT 10					
LOCUS	AF045450	40205 bp	DNA	linear	PRI 20-MAR-1998
DEFINITION	Homo sapiens chromosome 21q22.3 cosmid Q11M15, complete sequence.				
ACCESSION	AF045450				
VERSION	AF045450.1	GI:2895783			
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 40205)
JOURNAL	Taudien,S. and Rosenthal,A.
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 40205)
TITLE	Taudien,S., Nordstiek,G., Dagaad,E., Hildmann,T., Drescher,B., Weber,J., Rosenthal,A. and Yaspo,M.L.
JOURNAL	Direct Submission
FEATURES	Submitted (29-JAN-1998) Genome Analysis, Institut for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
SOURCE	Location/Qualifiers 1. .40205 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="21" /map="21q22.3" /clone="cosmid.Q11M15" complement (293. .586) /evidence=not experimental /rpt_family="AluSc" complement (643. .964) /evidence=not experimental /rpt_family="MERVL" complement (1053. .1176) /note="Genscan, score = 5.19%, comment = Internal_exon 124 bp frame: 2 phase: 1" /evidence=not experimental complement (1251. .1859) /evidence=not experimental /rpt_family="HERVL" complement (2296. .3320) /evidence=not experimental /rpt_family="HERV16" 2818. .2889 /note="Xpound exon prediction, score = 82% (0%)" /evidence=not experimental complement (3321. .3658) /evidence=not experimental /rpt_family="MLT1C" 3359. .3395 /note="Xpound exon prediction, score = 84% (0%)" /evidence=not experimental complement (3874. .4176) /evidence=not experimental /rpt_family="Aluub" complement (4330. .4433) /evidence=not experimental /rpt_family="L2" 4585. .4638 /note="GRAIL, score = 95.000%, comment = excellent shadow" /evidence=not experimental complement (5113. .7899) /evidence=not experimental /rpt_family="L1PA14" complement (6119. .6598) /note="GRAIL, score = 63.000%, comment = good" /evidence=not experimental complement (7937. .8045) /evidence=not experimental /rpt_family="FLAM_A" 8048. .8178 /evidence=not experimental /rpt_family="L2" 8771. .8876 /note="MZEF, score = 67.4%" /evidence=not experimental complement (8809. .9187) /evidence=not experimental /rpt_family="MERIA" 8837. .8876 /note="GRAIL, score = 55.000%, comment = good shadow" complement (9113. .9213)

/note="GRAIL, score = 55.000%, comment = good"
repeat_region /evidence=not experimental
complement(9150..9258)
/evidence=not experimental
/rpt_family="LIMC3"
9259..9647
repeat_region /evidence=not experimental
/rpt_family="LIMC3"
complement(10138..10427)
/evidence=not experimental
/rpt_family="MLT2E"
10662..10875
repeat_region /evidence=not experimental
/rpt_family="MER58A"
complement(11072..11156)
/note="GRAIL, score = 86.000%, comment = excellent"
exon /evidence=not experimental
complement(11072..11129)
/note="Xpound exon prediction, score = 89% (0%)"
repeat_region /evidence=not experimental
11478..11597
/evidence=not experimental
/rpt_family="L2"
11570..11704
exon /note="MZF, score = 66.1%"
/evidence=not experimental
11570..11704
/note="GRAIL, score = 99.000%, comment = excellent shadow"
repeat_region /evidence=not experimental
complement(11687..11735)
/evidence=not experimental
/rpt_family="MIR"
repeat_region /evidence=not experimental
complement(11774..12320)
/evidence=not experimental
/rpt_family="MLT1E"
exon /note="GRAIL, score = 11949"
complement(11828..11949)
/note="GRAIL, score = 50.000%, comment = good"
repeat_region /evidence=not experimental
12623..12843
/evidence=not experimental
/rpt_family="AlusX"
13102..13457
repeat_region /evidence=not experimental
/rpt_family="THB1B"
complement(13481..13519)
/note="GRAIL, score = 65.000%, comment = good"
repeat_region /evidence=not experimental
15065..15188
/evidence=not experimental
/rpt_family="HERVL"
15130..15378
misc_feature /note="GC score = 10.60 (249bp); Region: GC content"
repeat_region /evidence=not experimental
15369..16268
/evidence=not experimental
/rpt_family="MERVL"
exon /note="GRAIL, score = 16714"
complement(16623..16714)
/note="GRAIL, score = 72.000%, comment = good"
exon /evidence=not experimental
16661..16730
/note="MZF, score = 93.5%"
repeat_region /evidence=not experimental
16880..17318
/evidence=not experimental
/rpt_family="MLT1C"
17326..17436
repeat_region /evidence=not experimental
/rpt_family="LTR16C"
complement(17540..17675)
/evidence=not experimental
/rpt_family="MIR"
repeat_region complement(17805..17864)
/evidence=not experimental

/rpt_family="MLT1C"
exon /note="GRAIL, score = 45.000%, comment = marginal shadow"
18076..18098
/evidence=not experimental
complement(18219..18349)
/note="GenScan, score = 2.99%, comment = Internal_exon 131
bp frame: 1 phase: 2"
repeat_region /evidence=not experimental
complement(18717..19211)
/evidence=not experimental
/rpt_family="MLT1C"
exon complement(19252..19399)
/note="GRAIL, score = 71.000%, comment = good"
exon /evidence=not experimental
19269..19350
/note="GRAIL, score = 81.000%, comment = excellent shadow"
exon /evidence=not experimental
complement(19379..19399)
/note="Xpound exon prediction, score = 60% (0%)"
exon /evidence=not experimental
complement(19501..19881)
/note="MZF, score = 50.6%"
repeat_region /evidence=not experimental
19945..20020
/note="homology = 100.00%, score = 38, counts = 2"
/evidence=not experimental
/rpt_type=tandem
repeat_region /rpt_unit=tctctgcacattctctgtgagaggtcggttatcgc
20223..20643
/evidence=not experimental
/rpt_family="MLT1E"
repeat_region 20645..20910
/evidence=not experimental
/rpt_family="MLT1F"
repeat_region 20961..21007
/evidence=not experimental
/rpt_family="MLT1F"
repeat_region /evidence=not experimental
21483..21781
/evidence=not experimental
/rpt_family="AluDb"
exon complement(21504..21756)
/note="MZF, score = 79.4%"
repeat_region /evidence=not experimental
complement(22325..22806)
/evidence=not experimental
/rpt_family="MLT1C"

Query Match 86.8%; Score 435; DB 9; Length 40205;
Best Local Similarity 93.9%; Pred. No. 1.1e-118;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGATTACATCATGAGGCTGTGTTCTTATAGCCAGCTGACCTAGCCCT 60
|||
DB 30626 CAGATCTCAGATTACATCATGAGGCTGTGTTCTTATAGCCAGCTGACCTAGCCCT 30567
QY 61 TATACCTGCTTCCCAATATACAGAGAGAGAGAGGTTTACAGCTCGACCTCAG 120
|||
DB 30566 TATACCTGCTTCCCAATATACAGAGAGAGAGAGGTTTACAGCTCGACCTCAG 30507
QY 121 GATGCTTCTTGCATCCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAAGAT 180
|||
DB 30506 GATGCTTCTTGCATCCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAAGAT 30447
QY 181 ACTTCAAAACCAATCTTCACTCACTGACATATTTTACCCCAAGGTTCAAGGAATGT 240
|||
DB 30446 CTTTCAAAACCAATCTTCACTCACTGACATATTTTACCCCAAGGTTCAAGGAATGT 30387
QY 241 CCCCATATTTTGGCCAGGATTAGCCCAAGACTTGAGCCCAATCCCTACCTGACACT 300
|||
DB 30386 CCCCATATTTTGGCCAGGATTAGCCCAAGACTTGAGCCCAATCCCTACCTGACACT 30327
QY 301 -TGTCTTGGTGAAGTGAATTTACTTTTGGCCGCCCATTCAGAAACCTTGTGCAT 358
|||

Db	30326	CCTGTCCTTCAAGTCATGATGATATTTACTTTAGTGGCCCGTTTCAGAAACCTTGCGCAT	30267
QY	359	CAAGCCACCCCAAGCGCTCTTCAATTTCTCGCTA/CTGTGGCTACATGGTTTCCAAACCA	418
Db	30266	CAAGCCACCCCAAGCGCTTAAATTTCTCAGCTA/CTGTGGCTACAGAGTTTCCAAACCA	30207
QY	419	AAGGCTCAACTCTGCTCACAAGGTTA---CT..AGGGCTAAATTTATCCAAAGGACCC	474
Db	30206	AAGGCTCAACTCTGCTCACAAGGTTAAATGCT..AGGGCTAAATTTATCCAAAGTACCC	30147
QY	475	AGGGCCCTCAGTGAGGACACATCCAG	501
Db	30146	AGGGCCCTCAGTGAGGACATTCAG	30120
RESULT	11		
LOCUS	AF121782/c	142742 bp	DNA
DEFINITION	Homo sapiens chromosome 21q22.3	FAC 206A10, complete sequence.	
ACCESSION	AF121782		
VERSION	AF121782.1	GI:4210991	
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 142742)		
AUTHORS	Taudien,S., Dagdang,E., Hildmann,T., Nordstiek,G., Drescher,B., Schattervov,R., Weber,J., Schilling,M., Menzel,U., Yaspo,M.L. and Rosenthal,A.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-JAN-1999) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany		
FEATURES	Location/Qualifiers		
source	1..142742		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="21"		
	/map="21q22.3"		
	/clone="PAC 206A10"		
repeat_region	3..118		
	/evidence=not_experimental		
	/rpt_family="MIR"		
repeat_region	complement(140..415)		
	/evidence=not_experimental		
	/rpt_family="MLT2"		
repeat_region	complement(668..881)		
	/evidence=not_experimental		
	/rpt_family="MER58A"		
exon	complement(1078..1162)		
	/note="GRAIL, score = 86.000%, comment = excellent shadow"		
repeat_region	1114..1403		
	/evidence=not_experimental		
	/rpt_family="MLT1"		
repeat_region	1481..1603		
	/evidence=not_experimental		
	/rpt_family="L2"		
exon	1576..1710		
	/note="WZEF, score = 66.1%"		
	/evidence=not_experimental		
	/note="GRAIL, score = 99.000%, comment = excellent"		
exon	1576..1710		
	/note="GRAIL, score = 99.000%, comment = excellent"		
repeat_region	complement(1653..1741)		
	/evidence=not_experimental		
	/rpt_family="MIR"		
repeat_region	complement(1780..2326)		
	/evidence=not_experimental		
	/rpt_family="MLT1"		
exon	complement(1834..1955)		
	/note="GRAIL, score = 50.000%, comment = good shadow"		
	/evidence=not_experimental		
exon			
repeat_region	2629..2831		
	/evidence=not_experimental		
	/rpt_family="AluX"		
repeat_region	3012..3122		
	/evidence=not_experimental		
	/rpt_family="MLT1"		
repeat_region	3123..3478		
	/evidence=not_experimental		
	/rpt_family="THB1B"		
repeat_region	3479..3618		
	/evidence=not_experimental		
	/rpt_family="MLT1"		
exon	complement(3502..3540)		
	/note="GRAIL, score = 65.000%, comment = good shadow"		
	/evidence=not_experimental		
	/note="GRAIL, score = 42.000%, comment = marginal"		
exon	4674..4781		
	/evidence=not_experimental		
	/note="GRAIL, score = 42.000%, comment = marginal"		
	/evidence=not_experimental		
exon	4712..4749		
	/note="Xpound exon prediction, score = 70% (0%)"		
	/evidence=not_experimental		
repeat_region	5385..6104		
	/evidence=not_experimental		
	/rpt_family="HERV1"		
exon	6682..6751		
	/note="WZEF, score = 95.1%"		
	/evidence=not_experimental		
exon	6746..6751		
	/note="Genscan, score = 1.83%, comment = Initial_exon 6 bp frame: 1 phase: 0"		
	/evidence=not_experimental		
repeat_region	6901..7339		
	/evidence=not_experimental		
	/rpt_family="MLT1C"		
repeat_region	7347..7457		
	/evidence=not_experimental		
	/rpt_family="LTR16C"		
repeat_region	complement(7561..7696)		
	/evidence=not_experimental		
	/rpt_family="MIR"		
repeat_region	complement(7819..8220)		
	/evidence=not_experimental		
	/rpt_family="MLT1J"		
exon	8097..8119		
	/note="GRAIL, score = 45.000%, comment = marginal"		
	/evidence=not_experimental		
repeat_region	complement(8738..9232)		
	/evidence=not_experimental		
	/rpt_family="MLT1C"		
exon	complement(9273..9420)		
	/note="GRAIL, score = 71.000%, comment = good"		
	/evidence=not_experimental		
exon	9290..9371		
	/note="GRAIL, score = 81.000%, comment = excellent shadow"		
	/evidence=not_experimental		
repeat_region	9966..10041		
	/note="homology = 100.00%, score = 38, counts = 2"		
	/evidence=not_experimental		
	/rpt_type="tandem"		
repeat_region	10244..10664		

```

/ntc="MZF, score = 79.4%"
/evidence=not experimental
complement (12346, .12827)
repeat_region
/ntc family="MLT1C"
12802, .13180
/evidence=not experimental
/rpt family="Alusg"
complement (13057, .13161)
/ntc="MZF, score = 76.1%"
/evidence=not experimental
complement (13226, .13364)
/evidence=not experimental
/rpt family="L1MC2"
complement (13613, .13908)
/evidence=not experimental
/rpt family="Alusx"
complement (14317, .14614)
/evidence=not experimental
/rpt family="L1R37A"
complement (14962, .15036)
/ntc="MZF, score = 93%"
/evidence=not experimental
15255, .15299
/evidence=not experimental
/rpt family="L1PA6"
complement (15469, .15497)
/ntc="Xpound exon prediction, score = 62% (0%)"
16045, .16768
/evidence=not experimental
/evidence=not experimental
/rpt family="MER21A"
complement (16555, .16555)
/ntc="GRAIL, score = 54.000%, comment = good"
16823, .16850
/evidence=not experimental
/rpt family="MER34"
complement (16851, .17208)
/evidence=not experimental
/rpt family="THER1A"
17209, .17632
/evidence=not experimental
/rpt family="MER34"
complement (17633, .17669)
/evidence=not experimental
/rpt family="Alusg1"
complement (18004, .18334)
/evidence=not experimental
/rpt family="L1R17"
complement (18335, .18632)
/evidence=not experimental
/rpt family="HERV17"
18352, .18420
/ntc="GRAIL, score = 51.000%, comment = good shadow"
/evidence=not experimental
complement (18397, .18584)
/ntc="GRAIL, score = 66.000%, comment = good"
/evidence=not experimental
complement (18643, .24647)
/evidence=not experimental
/rpt family="HERV17"
complement (19228, .19361)
/ntc="MZF, score = 92.8%"
/evidence=not experimental
20900, .21135
/ntc="MZF, score = 51.4%"
/evidence=not experimental
21150, .21275
/ntc="Xpound exon prediction, score = 74% (0%)"
/evidence=not experimental
86.8%; Score 435; DB 9; Length 142742;

```

```

Best Local Similarity 93.9%; Pred. No. 1e-118;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

QY 1 CAAATCTCAGGATTAATCAATGAGGCGTGTCTCTATAGCAGCTGATCCGCT 60
Db 20632 CAAATCTCAGGATTAATCAATGAGGCGTGTCTCTATAGCAGCTGATCCGCT 20573
QY 61 TATACCTGCTTCCCAATATACAGAGAGAGGTTTACAGTCTGACCTTCAG 120
Db 20572 TATACCTGCTTCCCAATATACAGAGAGAGGTTTACAGTCTGACCTTCAG 20513
QY 121 GATGCTTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCT 180
Db 20512 GATGCTTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCT 20453
QY 181 ACTTCAAAACCAATCTCAATCACTGACATTTTACCCCAAGGTTCAAGGAT 240
Db 20452 CTTTCAAAACCAATCTCAATCACTGACATTTTACCCCAAGGTTCAAGGAT 20393
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAACATCTGACACT 300
Db 20392 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAACATCTGACACT 20333
QY 301 -TGTCTTGGTGAAGTGATGATTTACTTTTGGCCGCTTCAAGACCTTGTGCAT 358
Db 20332 CTTGTCTTGGTGAAGTGATGATTTACTTTTGGCCGCTTCAAGACCTTGTGCAT 20273
QY 359 CAAAGCCCAAGGCTCTTCAATTTCTGCTACCTGTGTGCTACATGTTTCCAAACCA 418
Db 20272 CAAAGCCCAAGGCTCTTCAATTTCTGCTACCTGTGTGCTACATGTTTCCAAACCA 20213
QY 419 AAGGCTCACTCTGCTCAGACAGGTTA---CTTAGGGCTAAATATTCAAAGGACC 474
Db 20212 AAGGCTCACTCTGCTCAGACAGGTTAATGCTTAGGGCTAAATATTCAAAGTACC 20153
QY 475 AGGGCCCTCAGTGAGAAACATCCAG 501
Db 20152 AGGGCCCTCAGTGAGAAACATCCAG 20126

RESULT 12
HS21C080/c 340000 bp DNA linear PRI 24-MAY-2000
LOCUS HS21C080 Homo sapiens chromosome 21 segment HS21C080.
DEFINITION AL163280 AP001735 BA000005
ACCESSION AL163280.2 GI:7717369
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 340000)
REFERENCE
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totsuki,Y., Choi,D.K., Seda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polly,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schilhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Ngamane,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Norisue,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Riesemann,L., Dagand,F., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaspo,M.L.
Direct Submissio
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center Human Genome Research
Group * Institute of Molecular Biotechnology Genome Analysis *
Keio University School of Medicine Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of

```

```
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: sakaki@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: shimizudmb-med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@bf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
Location/Qualifiers
1. 34000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1. 125946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P160L9, 5' partial"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF121897"
73410. 243533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P70124"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF064860"
224137. 3340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P206A10, 3' partial"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF11782"
286528. 3340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="BAC-291B3, 3' partial"
/clone_lib="BAC library"
/note="Accession No. AF064857"
1088. 1566
repeat_region
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2398. 2495
repeat_region
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TRANDEM
complement(2496. 2878)
/note="TBE1C"
/rpt_family="LTR/MaLR"

* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: sakaki@gsc.riken.go.jp
* URL: http://hgp.gsc.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Dept. of Molecular Biology, *
Tokyo 160-8582, Japan,
* e.mail: shimizudmb-med.keio.ac.jp
* URL: http://adenine.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@bf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Ihnestrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
Location/Qualifiers
1. 34000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1. 125946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P160L9, 5' partial"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF121897"
73410. 243533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P70124"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF064860"
224137. 3340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="P206A10, 3' partial"
/clone_lib="RPC11,3-5 PAC library"
/note="Accession No. AF11782"
286528. 3340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="BAC-291B3, 3' partial"
/clone_lib="BAC library"
/note="Accession No. AF064857"
1088. 1566
repeat_region
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2398. 2495
repeat_region
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TRANDEM
complement(2496. 2878)
/note="TBE1C"
/rpt_family="LTR/MaLR"

repeat_region
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
2879. 2931
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TRANDEM
2932. 3004
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TRANDEM
3005. 3074
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TRANDEM
3473. 19157
/gene="SH3BGR"
join(<3473. 3565,12649. 12678,15331. 15460,18739. 19157)
/gene="SH3BGR"
join(3473. 3565,12649. 12678,15331. 15460)
/partial
/gene="SH3BGR"
/note="Accession No. X93498"
/codon_start=1
/product="21-Glutamic Acid Rich protein 21-GARP"
/protein_id="CA890445.1"
/db_xref="GI:7717370"
/translacion="GSEKAEGETEAQKESSDVGNLPEAQKNEEFGTATEETE
IMEAGAGEAEAEETAECEPEDEDS"
3473. 3565
/gene="SH3BGR"
/number=4
5542. 5829
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(6887. 7067)
/note="L1MD3"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
complement(8482. 8762)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(8865. 8964)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
9131. 9424
/note="AluSg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(9723. 9882)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(9989. 10077)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
complement(10078. 10429)
/note="TBE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(10430. 10651)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
10980. 11000
/rpt_type=DISPERSED
/note="(TTTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TRANDEM
complement(11002. 11282)
/note="AluSg"
/rpt_family="SINE/Alu"
```

```

repeat_region
complement(11315..11556)
/note="AluUc"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
11702..11753
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
12649..12678
/gene="SH3BGR"

exon

repeat_region
complement(13115..13425)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(13598..13892)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14013..14131
/note="(CA)n"
/rpt_family="Simple_repeat"

```

Query Match	86.8%;	Score 435;	DB 9;	Length 340000;
Best Local Similarity	93.9%;	Pred. No. 1e-118;		
Matches 476;	Conservative	0;	Mismatches 25;	Indels 6;
				Gaps 2

QY	1	CAATCTCAGAGATTATCAAGAGAGCGTGTTCCTCTATATAGCAGGTACCTAGCCCT	60
Db	244792	CAGAGTCTCAGAGATTATCAAGAGAGCGTGTTCCTCTATATAGCAGGTACCTAGCCCT	244733
QY	61	TATACTCTGCTTTCCCAATATCAAGAGAAACAGAGTGTTTACATCTGACCTTCAG	120
Db	244732	TATACTCTGCTTTCCCAATATCAAGAGAAACAGAGTGTTTACATCTGACTTTAAG	244673
QY	121	GATGCTTCTCTGCAATCCCGTATCATCTGACCTCTCAATTCTGTGTGGCTTGAAGAT	180
Db	244672	GATGCTTCTCTGCAATCCCGTATCATCTGACCTCTCAATTCTGTGTGAATTTGAAGAT	244613
QY	181	ACTTCAAAACCCACATCTCAACTCACCTGACATATTTTAACCCCAAGGTTTCAGGGATAGT	240
Db	244612	CCTTCAAAACCCACATCTCAACTCACCTGACATGTTTTTAAACCCCAAGGGTTCAAGGATAGT	244553
QY	241	CCCCTATCTATTTGGCCAGAGCATTTAGCCCAAGACTTTAGGCCATCTATATCTGACACT	300
Db	244552	CCCCTATCTATTTGGCCAGAGCATTTAGCCCAAGACTTTAGGCCATCTATATCTGACACT	244493
QY	301	--TGTCCTCGATGATGATGATTTACTTTTGGCCGCCCATTCAGAAACCTTGTGCAT	358
Db	244492	CCTGTCTTTCAGTGCATGAGATGATTTACTTTTTCAGTCCCGTTCAAGAAACCTTGTGCAT	244433
QY	359	CNAACCAACCAAGGCTCTTCAATTTCTCTGCTACTGTGCTACATGTTTCCAAACCA	418
Db	244432	CNAACCAACCAAGGCTCTTCAATTTCTCTGCTACTGTGCTACCAAGGTTTCCAAACCA	244373
QY	419	AAGGCTCAACTCTCTCAACAGCAGGTATA---CTTAGGCGTAAATAATATCAAGGACAC	474
Db	244372	AAGGCTCAAGCTCTCTCAACAGCAGGATTAATGCTTAGGCGTAAATAATATCAAGGACAC	244313
QY	475	AGGGCCCTCACTGATGAGAAACATCTCAG	501
Db	244312	AGGGCCCTCACTGATGAGAAACGATATCAG	244286

RESULT	13
LOCUS	AL162912
DEFINITION	Human DNA sequence from clone RP6-198C4 on chromosome Xq13.3-21.2,
ACCESSION	AL162912
VERSION	AL162912.1
KEYWORDS	HTG.
GI:	7406722

SOURCE ORGANISM	REFERENCE
human.	1 (pages 1 to 8339)
Homo sapiens	
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	

FEATURES	Location/Qualifiers
source	1. .8339

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q13-3.21.2"
/clone="RP6-196C4"
/clone_1fb="RPCL-6"
1..5641
repeat_region
/note="HVR17 repeat: matches 368..5930 of consensus"
5635..5673
repeat_region
/note="HVR17 repeat: matches 7419..7457 of consensus"
5664..5955
repeat_region
/note="HVR17 repeat: matches 8244..8523 of consensus"
5966..6136
repeat_region
/note="L1R17 repeat: matches 1..239 of consensus"
6195..6280
repeat_region
/note="L1R17 repeat: matches 239..324 of consensus"
BASE COUNT      2649 a      1846 c      1726 g      2118 t
ORIGIN

```

Query Match 85.5%; Score 428.6; DB 9; Length 8339;
Best Local Similarity 93.1%; Pred. No. 9,8e-117;
Matches 472; Conservative 0; Mismatches 29; Indels 6; Gaps

QY	1	CAAATCTCAGGAATATACATAGAGCGCTGTTCCCTTAAGGCAGCTGACCTAGCCCT	60
Db	3668	CAAAATCTCAGGATATACAGTAGAGCGCTGTTCTCTCTATACCCAGGTGACTAGCCCT	3727
QY	61	TATCTCTGCTTTCCCAATAACAGAGAAAGCAGAGTGTTTACAGTCTTGACCTTTAG	120
Db	3728	TATACCTGTGCTTTCCCAATAACAGAGAAAGCAGAGTGTTTACAGTCTTGACCTTTAG	3787
QY	121	GATGCCCTTCTTCGATCCCTCTGACATCTCGACTCTCAATTCTTGTTGGCTTTGAAGAT	180

Db	3788	GATGCTTTTCTGTGATCCCTGTAATCTGACTCTCAATTCTTGTTGCCCTTGAGAT	3847
Qy	181	ACTTCAAAACCAACATCTCAACTCACTGACATATTTTACCACAAGGTCAGGGATAGT	240
Db	3848	CCTTCAAAACCCAGACTCTCAACTCACTGAGACTTTTATACCACAAGGTTAAGGATAGC	3907
Qy	241	CCCCATCTATTTGGCCAGGACATTTAGCCCAAGACTTGAGCCAACTCTCATTACCTGACA	298
Db	3908	CCCCATCTATTTGGCCAGGACATTTAGCCCAAGACTTGAGCCAACTCTCATATGACACT	3967
Qy	299	CTTGCTCTTCGGTAGGTGATGATTTACTTTTGACCGCCATTGAGAAACCTTGGCCAT	358
Db	3968	CTTGCTCTTTGGTAGGTGATGATTTATTTTAAACCAACCATTCAGAAACCTTGGCCAT	4027
Qy	359	CAGGACACCCAAAGCCCTCTTCAATTTCCCTCGTACCTGGGCTATCATGTTCCAAACA	418
Db	4028	CAGGACACCCAAAGTCTCTTAAATTTCCCTCAGTACCTGGGCTTACAAAGTTTCCAAACA	4087
Qy	419	AAGGCTCAACTGTGCTACAGCAGGT---TACTTAGGGCTTAAATTTACCAAGGCACC	474
Db	4088	AAGGCTCAGCTGTGCTCACAAGGTTAAGTACTTAGGGCTTAAATTTATCCAAAGGCACC	4147
Qy	475	AGGGCCCTCAGTAGAGAAACACATCCAG	501
Db	4148	ACGGCCCTCAGTAGAGAAATGTATCCAG	4174

LOCUS	AX000964	3372 bp	DNA	linear	PAT 10-MAR-2000
DEFINITION	Sequence 9 from Patent WO902696.				
ACCESSION	AX000964				
VERSION	AX000964.1	GI:7241206			
KEYWORDS					
SOURCE	unidentified.				
ORGANISM	unidentified.				
REFERENCE	unclassified.				
AUTHORS	unclassified.				
TITLE	1 (bases 1 to 3372)				
JOURNAL	Beesme, F. and Blond, J. ENDONEURIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE DISEASES OR WITH PREGNANCY DISORDERS Patent: WO 902696-A 9 21-JAN-1999; BIO MERIEUX (FR); BESEME FREDERIC (FR)				
FEATURES	Location/Qualifiers				
source	1..3372				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
BASE COUNT	1047 a 835 c 711 g 773 t				
ORIGIN					
Query Match	85.5%; Score 428.4; DB 6; Length 3372;				
Best Local Similarity	99.8%; Pred. No. 1.2e-116;				
Matches 429; Conservative	0; Mismatches 1; Indels				
	Gaps				
Db	72 TTCCCAATACCGAGAGCAGAGTAGTGTTCACAG;CTTGAACCTTCAGAGTCCTTCTT	131			
Qy	4 TTCCCAATACCGAGAGCAGAGTAGTGTTCACAG;CTTGAACCTTCAGAGTCCTTCTT	63			
Qy	132 CTGCATCCCTGTACATCTGACTCTCAATCTTGTGTGCTTTGAAAGATACCTCAACC	191			
Db	64 CTGCATCCCTGTACATCTGACTCTCAATCTTGTGTGCTTTGAAAGATACCTCAACC	123			
Qy	192 AACATCTCAACTCACTCTGAGCTATTTTACCCCAAGGCTTCAGGATAGTCCCATCATTT	251			
Db	124 AACATCTCAACTCACTCTGAGCTATTTTACCCCAAGGCTTCAGGATAGTCCCATCATTT	183			
Qy	252 TGCCACGAGCTTAGCCCAAGACTTGAAGCAATCTCTACCTTGAGACACTTGCTCTTGAT	311			
Db	184 TGCCACGAGCTTAGCCCAAGACTTGAAGCAATCTCTACCTTGAGACACTTGCTCTTGAT	243			
Qy	312 AGGTGATGATTTACTTTTGCCGCGCCCATTCAGAACTTTGTGCTATCAAGCACCACAG	371			

Db	244	AGGAGGATGATTTCCTTTTGGCGGCCCAATTGAGAAACCTTGCGATTCAGACCA	303
QY	372	CGCTCTTCAATTTCTCGCTACCTGTGGCTACATGTTTCCAAACCAAGGCTCAACTCT	431
Db	304	CGCTTTCATTTTCTCGCTACCTGTGGCTACATGTTTCCAAACCAAGGCTCAACTCT	363
QY	432	GCTCAGCAGAGTTACTTAAAGGCTAAATATCCAAAGGACCAAGGCGCTCATGTAGGA	491
Db	364	GCTCAGCAGAGTTACTTAAAGGCTAAATATCCAAAGGACCAAGGCGCTCATGTAGGA	423
QY	492	ACAATCCAG 501	
Db	424	ACAATCCAG 433	

	LOCUS	AX027478	3372 bp	DNA	linear	PAT 16-SEP-2000
	DEFINITION	Sequence 28 from Patent FR2788784.				
	ACCESSION	AX027478				
	VERSION	AX027478.1	GI:10188442			
	KEYWORDS	human.				
	SOURCE	Homo sapiens				
	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
	REFERENCE	1 (bases 1 to 3372)				
	AUTHORS	Mallet,F., Voisset,C. and Paranhos,B.G.				
	JOURNAL	Patent: FR 2788784-A 28 28-JUL-2000;				
	FEATURES	Bio MERIEUX (FR)				
	source	Location/Qualifiers				
		1..3372				
		/organism="Homo sapiens"				
		/db_xref="taxon:9606"				
BASE COUNT	1047 a	835 c	711 g	779 t		
ORIGIN						
Query Match	85.5%	Score 428.4;	DB 6;	Length 3372;		
Best Local Similarity	99.8%;	Pred. No. 1.2e-116;				
Matches 429;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;		
QY	72	TTTCCCAAAATACCAGAGGAGAAGCAGAGTGCTTAACAAGTCCTGGACCTTCAGAGATGCCTTCTT	131			
Db	4	TTCGCCAATAATACAGAGAGANGAGCATGTGTTTACAGTCTCGAACCTTCAGAGATGCCTTCTT	63			
QY	132	CTGCATATCCCTGTACATCTCTACTCTCAATTCTTGTGCTTTGAAGATCTTCAAACCC	191			
Db	64	CTGCATATCCCTGTACATCTCTCAATCTTGTGCTTTGAAGATCTTCAAACCC	123			
QY	192	AACATCTCACTCACCTGGACATATTTTACCCCCAAGGTTCAAGGATAGTCCCACATCTAAT	251			
Db	124	AGCATCTCAACTCACCTGGACATATTTTACCCCCAAGGTTCAAGGATAGTCCCACATCTAAT	183			
QY	252	TGGCCAGGATTAGGCCAAGACCTTGAGCAATCTCATACCTGACACACTGTCTTCGGT	311			
Db	184	TGGCCAGGATTAGGCCAAGACCTTGAGCAATCTCATACCTGAGCACTGTCTTCGGT	243			
QY	312	AGGTGATGATTTACTTTTGGCCGCCCATTCAGAAAACCTTGTCATCAAGCCACCAAG	371			
Db	244	AGGTGATGATTTACTTTTGGCCGCCCATTCAGAAAACCTTGTCATCAAGCCACCAAG	303			
QY	372	CGETCTTCAATTTCTCGCTACCTGTGGGTACATAGTTTTCCAACCAAGGCTCAACTCT	431			
Db	304	CGETCTTCAATTTCTCGCTACCTGTGGGTACATAGTTTTCCAACCAAGGCTCAACTCT	363			
QY	432	GCCTCAGAGAGTTACTTGAAGGCTAAATTTATCCAAAGSCACAGGGCCCTCAGTGAGGA	491			
Db	364	GCTCAGAGAGTTACTTGAAGGCTAAATTTATCCAAAGSCACAGGGCCCTCAGTGAGGA	423			
QY	492	ACACATCCAG 501				
Db	424	ACACATCCAG 433				

Mon Apr 21 09:59:02 2003

us-09-719-554-3_copy_5000_5500.rge

Page 15

Search completed: April 19, 2003, 13:51:57
Job time : 1888.47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 19, 2003, 10:05:38 ; Search time 180.12 Seconds
(without alignments)
6263.892 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501
Sequence: 1 caagatctcagatataca.....tcagtggagacacatccag 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	501	100.0	10499	21	ABN97929	Human retroviral s	
2	501	100.0	56093	24	ABL61744	ColoN adenocarcin	
3	474	94.6	2938	20	AA525656	Human endogenous r	
4	474	94.6	2938	21	AA59206	Gag and partial p	
5	474	94.6	7466	23	AA568626	DNA encoding novel	
6	453	8	90.6	3831	23	AA57127	DNA encoding novel
7	442	88.2	7582	20	AA525655	DNA encoding novel	
8	442	88.2	7582	21	AA59215	Complete human end	
9	435	86.8	3867	23	AA568621	Human endogenous r	
						DNA encoding novel	

10	435	86.8	3867	23	AA576464	DNA encoding novel
11	435	86.8	3867	23	AA580471	DNA encoding novel
12	435	86.8	3867	23	AA576475	DNA encoding novel
13	435	86.8	4535	23	AA576205	DNA encoding novel
14	435	86.8	5154	23	AA567609	DNA encoding novel
15	428.4	85.5	3372	20	AA525663	Human endogenous r
16	428.4	85.5	3372	21	AA59213	Partial pol gene a
17	422.2	84.3	900	23	AA567599	DNA encoding novel
18	422.2	84.3	900	23	AA572228	DNA encoding novel
19	422.2	84.3	900	23	AA573911	DNA encoding novel
20	422.2	84.3	900	23	AA576470	DNA encoding novel
21	422.2	84.3	1493	23	AA567606	DNA encoding novel
22	422.2	84.3	1687	23	AA572234	DNA encoding novel
23	422.2	84.3	1729	23	AA577018	DNA encoding novel
24	422.2	84.3	1743	23	AA584207	DNA encoding novel
25	422.2	84.3	2304	19	AA543199	Multiple sclerosis
26	420.6	84.0	924	23	AA572224	DNA encoding novel
27	420.6	84.0	1660	23	AA573608	DNA encoding novel
28	420.6	84.0	1660	23	AA576459	DNA encoding novel
29	420.6	84.0	1660	23	AA584201	DNA encoding novel
30	420.6	84.0	1810	23	AA585801	DNA encoding novel
31	420.6	84.0	1810	23	AA586023	DNA encoding novel
32	420.6	84.0	1846	23	AA573605	DNA encoding novel
33	420.6	84.0	1846	23	AA584192	DNA encoding novel
34	420.6	84.0	2220	23	AA567604	DNA encoding novel
35	420.6	84.0	2220	23	AA573919	DNA encoding novel
36	420.6	84.0	2220	23	AA576473	DNA encoding novel
37	419.6	83.8	1158	16	AA5702815	Multiple sclerosis
38	419.6	83.8	1158	17	AA573773	Partial sequence o
39	419.6	83.8	1158	18	AA576527	MSRV-1B pol* seque
40	419.6	83.8	1158	19	AA543115	Multiple sclerosis
41	419.6	83.8	2365	19	AA543200	Multiple sclerosis
42	419.6	83.8	2391	19	AA576478	MSRV-1 pol gene se
43	419.6	83.8	2391	19	AA543166	Multiple sclerosis
44	419	83.6	1117	23	AA576460	DNA encoding novel
45	419	83.6	1389	23	AA567594	DNA encoding novel

ALIGNMENTS

RESULT 1
ID ABN97929 standard; DNA; 10499 BP.
XX
AC ABN97929;
XX
DT 01-AUG-2002 (first entry)
XX
DE Human retroviral sequence HERV-7g.
XX
KW Autoimmune disease; HERV-7g; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
XX
OS Human retrovirus.
XX
PN WO9967395-A1.
XX
PD 29-DEC-1999.
XX
PF 23-JUN-1999; 99WO-FR01513.
XX
PR 23-JUN-1998; 98FR-0007920.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Alliel PM, Perin J, Rieger F;
XX
DR WPI; 2000-160587/14.
XX
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7g,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -

XX Claim 3; Fig 1; 225bp; French.
XX
CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;
Query Match 100.0%; Score 501; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 1,6e-151;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTTATAGCCAGCTGTACTAGCCCT 60
DB 5000 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTTATAGCCAGCTGTACTAGCCCT 5059
QY 61 TATATCTGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 5060 TATATCTGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
QY 121 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 5120 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5179
QY 191 ACTTCAAAACCAATCTCACTGACTGATCTTATTTTACCAGGTTGAGGATAGT 240
DB 5180 ACTTCAAAACCAATCTCACTGACTGATCTTATTTTACCAGGTTGAGGATAGT 5239
QY 241 CCCCACTATTGTCGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 5240 CCCCACTATTGTCGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 5299
QY 301 TGTCTTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 5300 TGTCTTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5359
QY 361 AGCCACCAAGGCTCTTCAATTTCTGCTACCTTGTGCTACATGTTTCCAAACCAA 420
DB 5360 AGCCACCAAGGCTCTTCAATTTCTGCTACCTTGTGCTACATGTTTCCAAACCAA 5419
QY 421 GGCTCAACTGTCTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 5420 GGCTCAACTGTCTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5479
QY 481 CTCAGTGAAGAACATCCAG 501
DB 5480 CTCAGTGAAGAACATCCAG 5500

XX gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-23133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237315P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Sopet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set
XX
XX Claim 1; SEQ ID 81; 44bp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic

CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.

XX Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16681 T; 0 other;

XX Query Match 100.0%; Score 501; DB 24; Length 56093;

XX Best Local Similarity 100.0%; Pred. No. 3.4e-151;

XX Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGGATTTCAATGAGGCTGTTGCTCTATAGCCAGCTGACCTT 60
DB 33000 CAAGATCTCAGGATTTCAATGAGGCTGTTGCTCTATAGCCAGCTGACCTT 33059
QY 61 TATACCTGCTTCCCAATACGAGAGGAGGAGGTTTAAAGTCTGAGACCTT 120
DB 33060 TATACCTGCTTCCCAATACGAGAGGAGGAGGTTTAAAGTCTGAGACCTT 33119
QY 121 GATGCTTCTTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 180
DB 33120 GATGCTTCTTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 33179
QY 181 ACTTCAAAACCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAAT 240
DB 33180 ACTTCAAAACCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAAT 33239
QY 241 CCCCATCTATTTGGCCGAGGAGTTAGCCAGCTTGAAGCCATCTCTGAGACT 300
DB 33240 CCCCATCTATTTGGCCGAGGAGTTAGCCAGCTTGAAGCCATCTCTGAGACT 33299
QY 301 TGTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 33300 TGTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 33359
QY 361 AGCCACCAAGCGCTCTCAATTTCTCTGATCTGATCTGATCTGATCTGAT 420
DB 33360 AGCCACCAAGCGCTCTCAATTTCTCTGATCTGATCTGATCTGATCTGAT 33419
QY 421 GGCTCACTCTGCTCAAGCAGGTTACTTAGGCTAAATATCAAAAGGACAGG 480
DB 33420 GGCTCACTCTGCTCAAGCAGGTTACTTAGGCTAAATATCAAAAGGACAGG 33479
QY 481 CTCAGTGAGGAACATCCAG 501
DB 33480 CTCAGTGAGGAACATCCAG 33500

RESULT 3

AAK25656 AAK25656 standard; cDNA to mRNA; 2938 BP.

XX AAK25656;

XX 21-MAY-1999 (first entry)

XX Human endogenous retrovirus W clone cl.6A1.

XX Clone; human endogenous retrovirus; genome; autoimmune disease;

XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;

XX disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

XX Human endogenous retrovirus.

XX WO9902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-FR01442.

XX 07-JUL-1997; 97FR-0008815.

XX (INMR) BIO MERIEUX.

XX Beseme F, Blond JL, Boulton O, Mallet F, Mandrand B;

XX WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -

XX expressed exclusively in placenta and useful in diagnosis and

XX therapy of autoimmune disease, and abnormal or failed pregnancy

XX Claim 1; Page 49-52; 106pp; French.

XX This sequence represents clone cl.6A1 of the human endogenous retrovirus

XX (HERV) W genome. The nucleic acids, their fragments or peptides encoded

XX by them are markers of autoimmune disease (e.g. multiple sclerosis,

XX rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-

XX dependent diabetes and related pathologies) and of abnormal or

XX unsuccessful pregnancy and can be used as chromosomal markers for

XX susceptibility to these conditions, or proximity markers of genes

XX associated with this susceptibility.

XX Sequence 2938 BP; 878 A; 720 C; 646 G; 692 T; 2 other;

XX Query Match 94.6%; Score 474; DB 20; Length 2938;

XX Best Local Similarity 97.8%; Pred. No. 5e-143;

XX Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 CAAGATCTCAGGATTTCAAT-GAGGCTGTTGCTCTATAGCCAGCTGACCTT 59
DB 2079 CAAGATCTCAGGATTTCAATGAGGCTGTTGCTCTATAGCCAGCTGACCTT 2138
QY 60 TTATACCTGCTTCTTCCCAATACGAGAGGAGGAGTGTTTTACAGTCTGACCTTCA 119
DB 2139 TTATACCTGCTTCTTCCCAATACGAGAGGAGGAGTGTTTTACAGTCTGACCTTCA 2198
QY 120 GATGCTTCTTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGAT 179
DB 2199 GATGCTTCTTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGAT 2258
QY 180 TACTTCAAAACCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAAT 229
DB 2259 TACTTCAAAACCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAAT 2318
QY 240 TCCCATCTATTTGGCCGAGGAGTTAGCCAGCTTGAAGCCATCTCTGAGACT 259
DB 2319 TCCCATCTATTTGGCCGAGGAGTTAGCCAGCTTGAAGCCATCTCTGAGACT 2378
QY 300 TGTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 359
DB 2379 TGTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2438
QY 360 AAGCCACCAAGCGCTCTCAATTTCTCTGATCTGATCTGATCTGATCTGAT 419
DB 2439 AAGCCACCAAGCGCTCTCAATTTCTCTGATCTGATCTGATCTGATCTGAT 2498
QY 420 AGGCTCACTCTGCTCAAGCAGGTTACTTAGGCTAAATATCAAAAGGACAGG 479
DB 2499 AGGCTCACTCTGCTCAAGCAGGTTACTTAGGCTAAATATCAAAAGGACAGG 2558
QY 480 CTCAGTGAGGAACATCCAG 501
DB 2559 CTCAGTGAGGAACATCCAG 2580

RESULT 4

AAA59206 AAA59206 standard; DNA; 2938 BP.

XX

AC AAA59206;
XX 07-NOV-2000 (first entry)
DE Gag and partial pol gene fragment of HERV-W from human genome.
XX
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KM gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX Homo sapiens.
XX MO200043521-A2.
XX
XX 27-JUL-2000.
XX
XX 21-JAN-2000; 2000WO-FR00144.
XX
XX 21-JAN-1999; 99FR-0000888.
XX
XX (INMR) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Mallet F, Volasset C;
PI WPI; 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
PT diagnosis of autoimmune disease and complications of pregnancy,
PT contains at least part of the gag gene -
XX
XX Disclosure; Page 43; 53pp; French.
XX
XX The present sequence represents an endogenous retroviral nucleic acid
CC fragment, which is associated with an autoimmune disease, and is
CC integrated into the human genome. The fragment is originally derived
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The
CC HERV-W retrovirus is associated with autoimmune disease, failure of
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or
CC proteins derived from it, are useful for diagnosis of autoimmune
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.
CC The nucleic acid fragments may also be used for in situ labelling of
CC isolated chromosomes, while the transcription product can be used to
CC study or monitor T cell proliferation in vitro.
XX
SQ Sequence 2938 BP; 878 A; 720 C; 646 G; 692 T; 2 other;
Query March 94.6%; Score 474; DB 21; Length 2938;
Best Local Similarity 97.8%; Pred. No. 5a-143;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 1 CAAAGTCTCAGAGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGCTGACCTAGCCC 59
DB 2079 CAAAGTCTCAGAGATTATCAATGAGGCGCGTGTCTTATATCCAGCTGATCAGGCC 2138
QY 60 TTATATCTGCTTTTCCAAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
DB 2139 TTATATCTGCTTTTCCAAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
QY 120 GGAGGCTCTTCTGAGACCCGCTGATCCGACTCTCAATCTTGTGCTTGAAGA 179
DB 2199 GGAAGGCTCTTCTGAGACCCGCTGATCCGACTCTCAATCTTGTGCTTGAAGA 2258
QY 180 TACTTCAACCAACATCTCAACTGACCTGACCTATTTTACCCCAAGGTTGAGGATAG 239
DB 2259 TACTTCAACCAACATCTCAACTGACCTGACCTGACCTTTTACCCCAAGGTTGAGGATAG 2318
QY 240 TCCCATCTATTTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 299
DB 2319 CCCCATCTATTTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2378
QY 300 TTGTCCTCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 359
DB 2379 TTGTCCTCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2438

QY 360 AAGCCACCAAGCGCTCTCAATTCTCTCGTACCTGTGCTACATGCTTTCCAAACCA 419
DB 2439 AAGCCACCAAGCGCTCTCAATTCTCTCGTACCTGTGCTACATGCTTTCCAAACCA 2498
QY 420 AGGCTCAACTCTGTCTCAGACAGGTTACTTAAAGGCTAAATTTCCAAAGGAGCAGGAGC 479
DB 2499 AGGCTCAACTCTGTCTCAGACAGGTTACTTAAAGGCTAAATTTCCAAAGGAGCAGGAGC 2558
QY 480 CCTCAGTAGAGAACACATCCAG 501
DB 2559 CCTCAGTAGAGAACACATCCAG 2580
RESULT 5
AAS68626
ID AAS68626 standard; cDNA; 7466 BP.
XX
AC AAS68626;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4430.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSB-) HYSBQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG04439.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 4430; 103bp; English.
PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensic, gene mapping, identification of mutations in
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 7466 BP, 2217 A, 1880 C, 1614 G, 1754 T, 1 other;

Query Match 94.6%; Score 474; DB 23; Length 7466;
Best Local Similarity 97.8%; Pred. No. 7, 6e-143;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

```
QY 1 CAAGATCTCAGATTATCAAT-GAGGCTGTGTCTCTATAGCCAGCTGACTGACCC 59
DB 6607 CAAGATCTCAGATTATCAATGAGGCGCTGTGTCTCTATAGCCAGCTGACTGACCC 6666
QY 60 TTATACCTGCTTCCCAATACAGAGGAAAGCAGAGTGTTCACCTGACCTTCA 119
DB 6667 TTATACCTGCTTCCCAATACAGAGGAAAGCAGAGTGTTCACCTGACCTTCA 6726
QY 120 GAGTGCCTTCTTGCATCCCTGTACATCTGACTCAATCTTGTGCTTTGAGA 179
DB 6727 GAGTGCCTTCTTGCATCCCTGTACATCTGACTCAATCTTGTGCTTTGAGA 6786
QY 180 TACTTCAAAACCAACATCTCACTGACCTGACTATTTTAAACCCCAAGGGTTCA 239
DB 6787 TACTTCAAAACCAACATCTCACTGACCTGACTGTTTAAACCCCAAGGGTTCA 6846
QY 240 TCCCACTCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCACTCTGACAC 299
DB 6847 TCCCACTCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCACTCTGACAC 6906
QY 300 TTGTCTCTGAGTGTGATGATTAATTTTGGCCGCTTGAAGAAACCTTGACATC 359
DB 6907 TTGTCTCTGAGTGTGATGATTAATTTTGGCCGCTTGAAGAAACCTTGACATC 6966
QY 360 AAGCCACCAAGGCGCTTCAATTTCTGCTACCTGTGCTACATGCTTCCAAACCA 419
DB 6967 AAGCCACCAAGGCGCTTCAATTTCTGCTACCTGTGCTACATGCTTCCAAACCA 7026
QY 420 AGGCTCAACTCTGCTCAGCAGAGTTACTTAAAGGCTTAAATTCAAAGCACAGG 479
DB 7027 AGGCTCAACTCTGCTCAGCAGAGTTACTTAAAGGCTTAAATTCAAAGCACAGG 7086
QY 480 CCTCAGTAGGAACACATCCAG 501
DB 7087 CCTCAGTAGGAACACATCCAG 7108
```

RESULT 6

AA571727
ID AA571727 standard; cDNA, 3831 BP.

XX AA571727;

AC AA571727;

DE 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #7531.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001, 2001WO-US08631.

PR 31-MAR-2000, 2000US-0540217.

PR 23-AUG-2000, 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX MPI, 2001-639362/73.

DR P-PSDB; ABG07540.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 1; SEQ ID No 7531; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AA564197-AA594564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3831 BP, 1173 A, 953 C, 815 G, 890 T, 0 other;

Query Match 90.6%; Score 453.8; DB 23; Length 3831;
Best Local Similarity 95.2%; Pred. No. 2e-136;
Matches 479; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

```
QY 1 CAAGATCTCAGATTATCAATGAGGCTGTGTCTCTATAGCCAGCTGACTGACCT 60
DB 412 CAAGATCTCAGATTATCAATGAGGCTGTGTCTCTATAGCCAGCTGACTGACCT 471
QY 61 TATACCTGCTTCCCAATATCCAGAGAAAGAGTGTTCAGATCTGACCTTCAG 120
DB 472 TATACCTGCTTCCCAATATCCAGAGAAAGAGTGTTCAGATCTGACCTTCAG 531
QY 121 GATGCTTCTTGCATCCCTGTACATCTGACTCAATCTTGTGCTTGAAGAT 180
DB 532 GATGCTTCTTGCATCCCTGTACATCTGACTCAATCTTGTGCTTGAAGAT 591
QY 181 ACTTCAAAACCAACATCTCACTGAGTGAATTTTAAACCCCAAGGTTCAAGGAT 240
DB 592 CTTTCAAAACCAATGCTCACTGAGTGAATTTTAAACCCCAAGGTTCAAGGAT 651
QY 241 CCCATCTATTTGGCCAGGATTAAGCCCAAGACTTGAAGCACTCTCATACCTG 298
DB 652 CTTCACTATTTGGCCAGGATTAAGCCCAAGACTTGAAGCACTCTCATACCTG 711
QY 299 CTTGTCTCTGAGTGTGATGATTAATTTTGGCCGCTTCAAAAACCTTGTGCA 358
DB 712 CTTGTCTCTGAGTGTGATGATTAATTTTGGCCGCTTCAAAAACCTTGTGCA 771
QY 359 CAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCAACCA 418
DB 772 CAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCAACCA 831
QY 419 AAGGCTCAACTCTGCTCAGCAGGTTACTTAAAGGCTTAAATTCAAAGGCAAG 478
DB 832 AAGGCTCAACTCTGCTCAGCAGGTTACTTAAAGGCTTAAATTCAAAGGCAAG 891
QY 479 CCTCAGTAGGAACACATCCAG 501
DB 892 CCTCAGTAGGAACACATCCAG 914
```


PT	contains at least part of the gag gene
XX	
PS	Disclosure; Page 49-52; 53pp; French.
CC	The present sequence represents an endogenous retrovirus, which is associated with an autoimmune disease, and is integrated into the human genome. The retrovirus is human endogenous retrovirus W (HERV-W). The HERV-W retrovirus is associated with autoimmune disease, failure of pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or proteins derived from it, are useful for diagnosis of autoimmune disease (specifically multiple sclerosis) and for monitoring pregnancy. The nucleic acid fragments may also be used for in situ labelling of isolated chromosomes, while the transcription product can be used to study or monitor T cell proliferation in vitro.
XX	
SQ	Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 216 other;
	Query Match 88.2%; Score 442; DB 21; Length 7582;
	Best Local Similarity 91.3%; Pred. No. 1.8e-132;
	Matches 464; Conservative 31; Mismatches 6; Indels 7; Gaps 3
QY	1 CAAGATCTCAGGATTATTCAT-GAGCGTGTTCCTCTATAGCCAGCTGTAACCCC 59
Db	2678 CAAGATCTCAGGATTATTCATGAGGCGCGTTGCTTTATACCAGCTGTACTTGCCC 2737
QY	60 TTATACCTGCTCTCCCAAAATACAGAGAGACAAGATGGTTTACAGTCTGAGACCTTCA 119
Db	2738 TTATACCTGTMWTTCCCAAATACAGAGAGAACAGATGGTTTACASTCTGGAACCTTWA 2797
QY	120 GGATGCCCTTTCTTGCGATCCCTGTACATCCCTGATCATCTCAATCTGTGTGGCTTTGAAGA 179
Db	2798 GGATGCCCTTTCTTGCGATCCCTGTACATCCCTGATCATCTCAATCTGTGTGGCTTTGAAMA 2857
QY	180 TACTTCAACCAACCAATCTCACTCACTGACATAATTTTACCCCAAGSGTTCAGGAGTAG 239
Db	2858 TACTTCAACCAACCAATCTCACTCACTGACATRTTTTACCCCAAGSITTCAGGAGTAG 2917
QY	240 TCCCCATCTAATTTGGCCAGGACATTAGCCCAAGCTTGAGCCAATCCTCATACCTTGACA- 298
Db	2918 YCCCATCTAATTTGGCCAGGACATTAGCCCAAGCTTGAGCAARTWTTCATACCTTGACAC 2977
QY	299 -CTTGTCTTGGGTAGTGTGATTTACTTTTGGCGCCGCATTCAGAAAACCTTGCGCA 357
Db	2978 TCTTGTCTTGTGKMGKAGGATGATTTACTTTTGGCGCGCTTTCAGAAAACCTTGCGCA 3037
QY	358 TCAAGCACCAAGGAGCTTCAATTTCTGCTGCTACCTGAGCTCAGTGTTCACCAACC 417
Db	3038 TCAAGCACCAAGGAGCTTCAATTTCTGCTGCTACCTGAGCTCAGTGTTCACCAACS 3097
QY	418 AAAGGCTCAACTCTGCTCAGCAGAGGT----TACTTTAGGGCTAAATTTCCAAAGGCAC 473
Db	3098 ARARECTCARCTCTGCTCAGCAGAGGTAAATTTAGTGRCTRAARATTTCCAAAGGCAC 3157
QY	474 CAGGCGCTCAGTGAAGAACACATCCAG 501
Db	3158 CARGGCGCTCAGTGAAGAAAYRATCCAG 3185
RESULT 9	
AAS68621	AAS68621 standard; cDNA; 3867 BP.
XX	AAS68621;
DT	13-FEB-2002 (first entry)
DE	DNA encoding novel human diagnostic protein #4425.
KM	Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	Homo sapiens.

Pt		MO200175067-A2.
PD		11-OCT-2001.
XX		
PF		30-MAR-2001; 2001WO-US08631.
XX		
PR		31-MAR-2000; 2000US-0540217.
PR		23-AUG-2000; 2000US-0649167.
PA		(HYSE-) HYSEQ INC.
PI		Drmnac RT, Liu C, Tang YT;
XX		
DR		WPI; 2001-639362/73.
XX		P-PsDB; ABG04434.
PT		New isolated polynucleotide and encoded polypeptides, useful in
PT		diagnostics, forensics, gene mapping, identification of mutations
PT		responsible for genetic disorders or other traits and to assess
PT		biodiversity -
XX		
PS		Claim 1; SEQ ID No 4425; 103bp; English.
XX		
CC		The invention relates to isolated polynucleotide (I) and
CC		polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC		polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC		and gene mapping, and in recombinant production of (II). The
CC		polynucleotides are also used in diagnostics as expressed sequence tags
CC		for identifying expressed genes. (I) is useful in gene therapy techniques
CC		to restore normal activity of (II) or to treat disease states involving
CC		(II). (II) is useful for generating antibodies against it, detecting or
CC		quantitating a polypeptide in tissue, as molecular weight markers and as
CC		a food supplement. (II) and its binding partners are useful in medical
CC		imaging of sites expressing (II). (I) and (II) are useful for treating
CC		disorders involving aberrant protein expression or biological activity.
CC		The polypeptide and polynucleotide sequences have applications in
CC		diagnostics, forensics, gene mapping, identification of mutations
CC		responsible for genetic disorders or other traits to assess biodiversity
CC		and to produce other types of data and products dependent on DNA and
CC		amino acid sequences. AAS64197-AAS94564 represent novel human
CC		diagnostic coding sequences of the invention.
CC		Note: The sequence data for this patent did not appear in the printed
CC		specification, but was obtained in electronic format directly from WIPO
CC		at ftp.wipo.int/pub/published_pct_sequences.
XX		
SQ		Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;
		Query Match 86.8%; Score 435; DB 23; Length 3867;
		Best Local Similarity 93.9%; Pred. No. 2.5e-130;
		Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
OY	1	CAAGATCAGGATTATCATGAGCGCTTTGTTCCTCATAGCCAGCTGACTTACGCCCT 60
Db	1693	CAGAAGCTCAGGATTATCAATGAGGGCTGTCTTTCTCATACCCGCTATATCTTAGCCCT 1752
OY	61	TATACTTGCTTTCCCAATATACAGAGAGAGAGAGTGGTTTACAGTCTTGACCTTGAG 120
Db	1753	TATATCTGCTTTCCCAATATACAGAGAGAGAGAGTGGTTTACAGTCTTGATCTTAAAG 1812
OY	121	GATGCTTCTTCTGCATCCCTGTATACCTCTGACTCTCAATTTCTTTGGCTTTGAAGT 180
Db	1813	GATGCTTCTTCTGCATCCCTGTATACCTCTGACTCTCAATTTCTTTGGCATTTGAAGT 1872
OY	181	ACTTAAACCACAATCTCACTCACCTGACTATTTTAACCCCAAGGGTTACGGATAGT 240
Db	1873	CTTTAAACCACAATCTCACTCACCTGACTATTTTAACCCCAAGGGTTACGGATAGT 1932
OY	241	CCCCATATTTGGCCAGAGCAATTAGCCCAAGACTGAGCAATCTCTCATPACTTGAGACT 300
Db	1933	CCCCATATTTGGCCAGAGCAATTAGCCCAAGACTGAGCAATCTCTCATPACTTGAGACT 1992
OY	301	--TGTCTTGGTATGATGATGATTTACTTTTGGCCGCCCATTCAGAACTTGTGCAT 358

Db 1993 CCTGTCTCTTCAATGATGATGATTTACTTTAG:ATGCCCTTCAAGAACTTGCCAT 2052
Qy 359 CAAGCCACCCAGGCTCTTCAATTTCTCGCTA:CTGGGCTACATGTTCCAAACCA 418
Db 2053 CAAGCCACCCAGGCTCTTCAATTTCTCGCTA:CTGGGCTACATGTTCCAAACCA 2112
Qy 419 AAGGCTCAACTCTGCTCAACAGAGTTA----CTTAGGCTTAAATTTCCAAAGGACCC 474
Db 2113 AAGGCTCAAGCTCTGCTCAACAGAGTTAAATGCTTAGGCTTAAATTTCCAAAGTCAAC 2172
Qy 475 AGGGCCCTCAGTAGGAACACATCCAG 501
Db 2173 AGGGCCCTCAGTAGGAACGATCCAG 2199

RESULT 10
AAS76464
ID AAS76464 standard; cDNA; 3867 BP.
XX
AC AAS76464;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #12268.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI
DR WPI: 2001-639362/73.
DR P-PSDB; ABG12277.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 12268; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques for restoring normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations in responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO

CC at fcp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;
XX
Query Match 86.8%; Score 435; DB 23; Length 3867;
Best Local Similarity 93.9%; Pred. No. 2.5e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
Qy 1 CAAGATCTCAGATATTAATCAATGAGGCTGTGTTCTTATATAGCCAGCTGTAGCCCT 60
Db 1693 CAATATCTCAGATATTAATCAATGAGGCTGTGTTCTTATATAGCCAGCTGTAGCCCT 1752
Qy 61 TATATCTGCTTTTCCAAATACCAAGAGACAGAGGTTTACAGTCTTGACCTTCAAG 120
Db 1753 TATATCTGCTTTTCCAAATACCAAGAGACAGAGGTTTACAGTCTTGACCTTCAAG 1812
Qy 121 GATGCTTCTTCTGCAATCCCGTACATCTGACTCTCAATTTCTGTTGCTTGAAGAT 180
Db 1813 GATGCTTCTTCTGCAATCCCGTACATCTGACTCTCAATTTCTGTTGCTTGAAGAT 1872
Qy 181 ACTTCAAAACCAACATCTCAATCAGCTGACATATTTTAAACCCCAAGGTTCAAGGATAGT 240
Db 1873 CCTTCAAAACCAACATCTCAATCAGCTGACATATTTTAAACCCCAAGGTTCAAGGATAGT 1932
Qy 241 CCCCATCTATTGGCCCAAGCATTAAGCCCAAGACTTGAAGCCATCTTCAATCTGACACT 300
Db 1933 CCCCATCTATTGGCCCAAGCATTAAGCCCAAGACTTGAAGCCATCTTCAATCTGACACT 1992
Qy 301 --TGTCCTTGGTAGAGGATTTACTTTTGGCCGCCCATTCAGAAACCTTGCCAT 358
Db 1993 CCTGTCTCTTCAATGATGATGATTTACTTTTACTGCTGCCCTTCAAGAACTTTGCCAT 2052
Qy 359 CAAGCCACCCAGGCTCTTCAATTTCTCGCTA:CTGGGCTACATGTTCCAAACCA 418
Db 2053 CAAGCCACCCAGGCTCTTCAATTTCTCGCTA:CTGGGCTACATGTTCCAAACCA 2112
Qy 419 AAGGCTCAACTCTGCTCAACAGAGTTA----CTTAGGCTTAAATTTCCAAAGGACCC 474
Db 2113 AAGGCTCAAGCTCTGCTCAACAGAGTTAAATGCTTAGGCTTAAATTTCCAAAGTCAAC 2172
Qy 475 AGGGCCCTCAGTAGGAACACATCCAG 501
Db 2173 AGGGCCCTCAGTAGGAACGATCCAG 2199

RESULT 11
AAS80471
ID AAS80471 standard; cDNA; 3867 BP.
XX
AC AAS80471;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #16275.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI
XX

DR WPI, 2001-639362/73.
XX P-PSDB; ABG16284.
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 1; SEQ ID No 16275; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;

Query Match 86.8%; Score 435; DB 23; Length 3867;
Best Local Similarity 93.9%; Pred. No. 2.5e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

QY 1 CAAGATCTCGAGTATTCATGAGAGGCTGCTGCTCATATGCGAGCTGATCCTAGCCCT 60
DB 1693 CAAGATCTCGAGTATTCATGAGAGGCTGCTGCTCATATGCGAGCTGATCCTAGCCCT 1752
QY 61 TATATCTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 1753 TATATCTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1812
QY 121 GATGCTCTTCTTGCATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 180
DB 1813 GATGCTCTTCTTGCATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 1872
QY 181 ACTTCAATCCCAATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 240
DB 1873 CTTTCAATCCCAATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 1932
QY 241 CCCCATCTATTTGGCCAGGAGGATGAGCCCAAGACTTGAGCCCAATCTGATCTGAGCACT 300
DB 1933 CCCCATCTATTTGGCCAGGAGGATGAGCCCAAGACTTGAGCCCAATCTGATCTGAGCACT 1992
QY 301 -TGTCTTGGGAGGAGGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGGCCAT 358
DB 1993 CTTGTCTTGGGAGGAGGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGGCCAT 2052
QY 359 CAAGCCACCCCAAGGCTCTTCAATTTCTGCTACCTGCTGCTGATCTGATCTGATCTGATCT 418
DB 2053 CAAGCCACCCCAAGGCTCTTCAATTTCTGCTACCTGCTGCTGATCTGATCTGATCTGATCT 2112
QY 419 AAGGCTCAACTCTGCTACAGCAGGATTA---CTTAAAGGCTTAAATTTCAAAAGGACCC 474
DB 2113 AAGGCTCAAGCTCTGCTACAGCAGGATTAAGCTTAAAGGCTTAAATTTCAAAAGTCAAC 2172
QY 475 AAGGCTCAAGCTCTGCTACAGCAGGATTAAGCTTAAAGGCTTAAATTTCAAAAGTCAAC 501
DB 2173 AAGGCTCAAGCTCTGCTACAGCAGGATTAAGCTTAAAGGCTTAAATTTCAAAAGTCAAC 2199

RESULT 12
AAS76475
ID AAS76475 standard; cDNA; 4349 BP.
XX
XX AAS76475;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #12279.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001MO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX
XX Dmanac RT, Liu C, Tang YT;
PT
XX
XX WPI; 2001-639362/73.
DR
XX
XX P-PSDB; ABG12288.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 1; SEQ ID No 12279; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4349 BP; 1319 A; 1061 C; 941 G; 1026 T; 2 other;

Query Match 86.8%; Score 435; DB 23; Length 4349;
Best Local Similarity 93.9%; Pred. No. 2.6e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

QY 1 CAAGATCTCGAGTATTCATGAGAGGCTGCTGCTCATATGCGAGCTGATCCTAGCCCT 60
DB 2382 CAAGATCTCGAGTATTCATGAGAGGCTGCTGCTCATATGCGAGCTGATCCTAGCCCT 2441
QY 61 TATATCTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

```
Db 2442 TATACCTGCTTTCCCAATACACAGAGAAAGAGTGGTTACAGTCTGATCTTAAG 2501
Qy 121 GATGCTCTCTCTGATGATCCGTCATCTGATCTCAATCTGTTGGCTTTGAAGAT 180
Db 2502 GATGCTCTCTCTGATGATCCGTCATCTGATCTCAATCTGTTGGCTTTGAAGAT 2561
Qy 181 ACTTCAACCAACATCTCACTCACTGACTATTTTACCACAGGGTTGAGGATAGT 240
Db 2562 CCTTCAACCAACATCTCACTCACTGACTATTTTACCACAGGGTTGAGGATAGT 2621
Qy 241 CCCCATCTATTTGGCCAGGACATTAGCCCAAGACTTGAGCCATCTCTATCTGGACACT 300
Db 2622 CCCCATCTATTTGGCCAGGACATTAGCCCAAGACTTGAGCCATCTCTATCTGGACACT 2681
Qy 301 --TGCCCTCGGTAGTGGATGATTTACTTTGG--CGCCCAATGCAAGAACTTGCCCAT 358
Db 2682 CCGTCTCTTCAGTGCATGATGATTTACTTTAG--TGCCCGTTTCAGAAACCTTGTCAT 2741
Qy 359 CAAGCCACCCCAAGCGCTCTTCAATTTCTCGCTA--CTGTGCTACATGATGTTTCCAAACCA 418
Db 2742 CAAGCCACCCCAAGCGCTCTTCAATTTCTCGCTA--CTGTGCTACATGATGTTTCCAAACCA 2801
Qy 419 AAGGCTCACTCTGCTCAACAGAGTTA---CTTAAAGGCTTAAATTTTCCAAAGGACAC 474
Db 2802 AAGGCTCACTCTGCTCAACAGAGTTAATGCTTAAAGGCTTAAATTTTCCAAAGTCAAC 2861
Qy 475 AAGGCGCTCAGTGAAGAACATCCAG 501
Db 2862 AAGGCGCTCAGTGAAGAACATCCAG 2888
```

RESULT 13

```
AA576205
ID AA576205 standard; cDNA; 4535 BP.
XX
AC AA576205;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #12009.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN MO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR P-PSDB; ABG12018.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostic, forensic, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
PS Claim 1; SEQ ID No 12009; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
```

CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful for treating
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct_sequences.

XX Sequence 4535 BP; 1356 A; 1159 C; 983 G; 1037 T; 0 other;

Query Match 86.8%; Score 435; DB 23; Length 4535;

Best Local Similarity 93.9%; Pred. No. 2, 7e-130; Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

```
Qy 1 CAAGATCTCAGGATTTATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGTAACCTT 60
Db 1693 CAAGATCTCAGGATTTATCAATGAGGCTGTTTCTCTATAGCCAGCTATACCTAGCCCT 1752
Qy 61 TATATCTGCTTTTCCCAATATCCAGAGAAAGAGTTCATGATCCAGACTTCAAG 120
Db 1753 TATATCTGCTTTTCCCAATATCCAGAGAAAGAGTTCATGATCCAGACTTCAAG 1812
Qy 121 GATGCTCTCTCTGATGATCCGTCATCTGATCTCAATCTGTTGGCTTTGAAGAT 180
Db 1813 GATGCTCTCTCTGATGATCCGTCATCTGATCTCAATCTGTTGGCTTTGAAGAT 1872
Qy 181 ACTTCAACCAACATCTCAATCTGATCTGATCTCAATCTGTTGGCTTTGAAGAT 240
Db 1873 CCTTCAACCAACATCTCAATCTGATCTGATCTCAATCTGTTGGCTTTGAAGAT 1932
Qy 241 CCCCATCTATTTGGCCAGGACATTAGCCCAAGACTTGAGCCCAATCTCTATCTGGACACT 300
Db 1933 CCCCATCTATTTGGCCAGGACATTAGCCCAAGACTTGAGCCCAATCTCTATCTGGACACT 1992
Qy 301 --TGCCCTCGGTAGTGGATGATTTACTTTTGGCCCGCCCATTCAGAAACCTTGCCAT 358
Db 1993 CCGTCTCTTCAGTGCATGATGATTTACTTTTGGCCCGCCCATTCAGAAACCTTGCCAT 2052
Qy 359 CAAGCCACCCCAAGCGCTCTTCAATTTCTCGCTA--CTGTGCTACATGATGTTTCCAAACCA 418
Db 2053 CAAGCCACCCCAAGCGCTCTTCAATTTCTCGCTA--CTGTGCTACATGATGTTTCCAAACCA 2112
Qy 419 AAGGCTCACTCTGCTCAACAGAGTTA---CTTAAAGGCTTAAATTTTCCAAAGGACAC 474
Db 2113 AAGGCTCACTCTGCTCAACAGAGTTAATGCTTAAAGGCTTAAATTTTCCAAAGTCAAC 2172
Qy 475 AAGGCGCTCAGTGAAGAACATCCAG 501
Db 2173 AAGGCGCTCAGTGAAGAACATCCAG 2199
```

RESULT 14

```
AA567609
ID AA567609 standard; cDNA; 5154 BP.
XX
AC AA567609;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #3413.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
```

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX MO200175067-A2.
 PN 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 XX P-PSDB; AEG03422.
 DR
 XX
 PT New isolated polynucleotide and encoded polypeptide, useful in
 PT diagnostics, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 1; SEQ ID No 3413; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (II) is useful as hybridisation probe,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensic, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 5154 BP; 1537 A; 1321 C; 1117 G; 1179 T; 0 other;
 SQ
 Query Match 86.8%; Score 435; DB 23; Length 5154;
 Best Local Similarity 93.9%; Pred. No. 2.9e-130;
 Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
 XX
 QY 1 CAAGATCTCAGATTTAATCAATGAGGCTGTGTTCTCTATAGCAGGCTGATACCTAGCCCT 60
 DB 1693 CAAGATCTCAGATTTAATCAATGAGGCTGTGTTCTCTATAGCAGGCTGATACCTAGCCCT 1752
 QY 61 TATATCTGCTTTCCAAATACAGAGGAAAGAGAGTGTGTTTACAGTCTTGAGACTTTTCAAG 120
 DB 1753 TATATCTGCTTTCCAAATACAGAGGAAAGAGAGTGTGTTTACAGTCTTGAGACTTTTCAAG 1812
 QY 121 GATGCTTTCTTGCAGATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 180
 DB 1813 GATGCTTTCTTGCAGATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 1872
 QY 181 ACTTCAACCAACATCTCAACCTGAGTACTATTTTACCCCAAGGCTTCAAGGATAGT 240
 DB 1873 CTTTCAACCAACATCTCAACCTGAGTACTATTTTACCCCAAGGCTTCAAGGATAGT 1932
 QY 241 CCCCATCTTATTTGGCCAGGACTTGAAGCCCAAGCTTGAGCCCAATCTTCACTGAGACT 300
 |||||||

DB 1933 CCCCATCTTATTTGGCCAGGACTTGAAGCCCAAGCTTGAGCCCAATCTTCACTGAGACT 1992
 QY 301 --TGTCTTCTGAGTGTGATGATTTACTTTGGCCCATTCAGAAACCTTGCCAT 358
 DB 1993 CTTGTCTTCTGAGTGTGATGATTTACTTTGGCCCATTCAGAAACCTTGCCAT 2052
 QY 359 CAAGCCACCCAGGCTCTTCAATTTCTGCTACTGTTGGCTTACATGTTTCCAAACCA 418
 DB 2053 CAAGCCACCCAGGCTCTTCAATTTCTGCTACTGTTGGCTTACATGTTTCCAAACCA 2112
 QY 419 AAGGCTCACTCTGCTCAGAGGCTTA---CTTAGGGCTAAATATTCAGAAAGCACC 474
 DB 2113 AAGGCTCACTCTGCTCAGAGGCTTAATGCTTAAATGCTTAAATATTCAGAAAGCACC 2172
 QY 475 AGGCGCTCAGTGAAGACATCCAG 501
 DB 2173 AGGCGCTCAGTGAAGACATCCAG 2199
 |||||||
 RESULT 15
 AAX25663
 ID AAX25663 standard; cDNA to mRNA; 3372 BP.
 XX
 AC AAX25663;
 XX
 DT 21-MAY-1999 (first entry)
 XX
 DE Human endogenous retrovirus W clone cl.P15T.
 KW
 KW clone; human endogenous retrovirus; genome; autoimmune disease;
 KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
 XX
 OS Human endogenous retrovirus.
 XX
 XX PN WO902656-A1.
 XX
 XX 21-JAN-1999.
 PD
 XX
 PF 06-JUL-1998; 98WO-FR01442.
 XX
 XX 07-JUL-1997; 97FR-0008815.
 PR
 XX
 PA (INMR) BIO MERIEUX.
 XX
 PI Besene F, Blond JL, Bouton O, Mallet F, Mandrand B;
 XX
 XX WPI; 1999-120897/10.
 DR
 XX
 PT New nucleic acid sequences from human endogenous retrovirus-W -
 PT expressed exclusively in placenta and useful in diagnosis and
 PT therapy of autoimmune disease, and abnormal or failed pregnancy
 PT
 XX
 PS Claim 1; Page 64-67; 106bp; French.
 XX
 CC This sequence represents clone cl.P15T of the human endogenous retrovirus
 CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded
 CC by them are markers of autoimmune disease (e.g. multiple sclerosis,
 CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
 CC dependent diabetes and related pathologies) and of abnormal or
 CC unsuccessful pregnancy and can be used as chromosomal markers for
 CC susceptibility to these conditions, or proximity markers of genes
 CC associated with this susceptibility.
 CC
 XX
 XX Sequence 3372 BP; 1047 A; 835 C; 711 G; 779 T; 0 other;
 SQ
 Query Match 85.5%; Score 428.4; DB 20; Length 3372;
 Best Local Similarity 99.8%; Pred. No. 3.3e-128;
 Matches 429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 XX
 QY 72 TTCCCAAAATACAGAGGAAAGAGTGTGTTTACAGTCTGAGACTTCAAGATCCCTTTCT 131
 DB 4 TTCCCAAAATACAGAGGAAAGAGTGTGTTTACAGTCTGAGACTTCAAGATCCCTTTCT 63
 |||||||

```
QY 132 CTGATCCCTGTAATGCTGACTCTCAATTCTGTGTTGCCCTTGAGATACTTCAAAACC 191
    |||
Db 64 CTGATCCCTGTAATGCTGACTCTCAATTCTGTGTTGCCCTTGAGATACTTCAAAACC 123
    |||
QY 192 AACATCTCACTCACTGACTATATTTTACCCCAAGGTTGAGGATAGTCCCATCTATT 251
    |||
Db 124 AGCATCTCACTCACTGACTATATTTTACCCCAAGGTTGAGGATAGTCCCATCTATT 183
    |||
QY 252 TGGCAGGCAATTAGCCCAAGACTTGGCCCAATCTCATACCTGAGCACTGTCTTGGT 311
    |||
Db 184 TGGCAGGCAATTAGCCCAAGACTTGGCCCAATCTCATACCTGAGCACTGTCTTGGT 243
    |||
QY 312 AGGTGATGATTTCTTTGGCCGCCCATTCAGAACTTGTGCTCAAGCCCAAG 371
    |||
Db 244 AGGTGATGATTTCTTTGGCCGCCCATTCAGAACTTGTGCTCAAGCCCAAG 303
    |||
QY 372 CGCTCTTCAATTTCTGCTACTGTGCTACATGTTTCCAAACCAAGGCTCACTCT 431
    |||
Db 304 CGCTCTTCAATTTCTGCTACTGTGCTACATGTTTCCAAACCAAGGCTCACTCT 363
    |||
QY 432 GCTCAGACAGGTTACTTAGGGCTTAAATTATCCAAAGGCAACGAGGCTTCACTGAGA 491
    |||
Db 364 GCTCAGACAGGTTACTTAGGGCTTAAATTATCCAAAGGCAACGAGGCTTCACTGAGA 423
    |||
QY 492 ACACATCCAG 501
    |||
Db 424 ACACATCCAG 433
    |||
```

Search completed: Apr:19, 2003, 12:18:09
Job time : 210.12 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 12:03:58 ; Search time 34.6897 Seconds
(without alignments)
4429.120 Million cell updates/sec

Title: US-09-719-554-3_COPY_5500_5500
Perfect score: 501
Sequence: 1 caagatccagcagatcatca.....tcagtgcagcagacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfillseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	419.6	83.8	1158	1 US-08-471-724-1	Sequence 1, Appl
2	419.6	83.8	1158	2 US-08-471-969-1	Sequence 1, Appl
3	419.6	83.8	1158	2 US-08-384-137-1	Sequence 1, Appl
4	419.6	83.8	1158	2 US-08-470-006A-1	Sequence 1, Appl
5	419.6	83.8	1158	3 US-08-691-563C-1	Sequence 1, Appl
6	419.6	83.8	1158	4 US-09-200-990-1	Sequence 1, Appl
7	419.6	83.8	1158	4 US-09-133-411-1	Sequence 1, Appl
8	419.6	83.8	2391	3 US-08-691-563C-57	Sequence 57, Appl
9	404.4	80.7	1577	3 US-08-691-563C-89	Sequence 89, Appl
10	271.8	54.3	2330	4 US-09-120-653D-4	Sequence 4, Appl
11	271.8	54.3	3910	4 US-09-120-653D-1	Sequence 1, Appl
12	212.6	42.4	297	1 US-08-471-724-2	Sequence 2, Appl
13	212.6	42.4	297	2 US-08-471-969-2	Sequence 2, Appl
14	212.6	42.4	297	2 US-08-384-137-2	Sequence 2, Appl
15	212.6	42.4	297	3 US-08-470-006A-2	Sequence 2, Appl
16	212.6	42.4	297	3 US-08-691-563C-2	Sequence 2, Appl
17	212.6	42.4	297	4 US-09-200-990-2	Sequence 2, Appl
18	212.6	42.4	297	4 US-09-133-411-2	Sequence 2, Appl
19	211.4	42.2	645	1 US-08-471-724-8	Sequence 8, Appl
20	211.4	42.2	645	1 US-08-471-969-8	Sequence 8, Appl
21	211.4	42.2	645	2 US-08-384-137-8	Sequence 8, Appl
22	211.4	42.2	645	2 US-08-470-006A-8	Sequence 8, Appl
23	211.4	42.2	645	3 US-08-691-563C-8	Sequence 8, Appl
24	211.4	42.2	645	4 US-09-200-990-8	Sequence 8, Appl
25	211.4	42.2	645	4 US-09-133-411-8	Sequence 8, Appl
26	128.8	25.7	299	3 US-08-691-563C-40	Sequence 40, Appl
27	116.4	23.2	2448	3 US-08-691-563C-53	Sequence 53, Appl

28	109.8	21.9	741	1 US-08-471-724-9	Sequence 9, Appl
29	109.8	21.9	741	2 US-08-471-969-9	Sequence 9, Appl
30	109.8	21.9	741	2 US-08-384-137-9	Sequence 9, Appl
31	109.8	21.9	741	2 US-08-470-006A-9	Sequence 9, Appl
32	109.8	21.9	741	3 US-08-691-563C-9	Sequence 9, Appl
33	109.8	21.9	741	4 US-09-200-990-9	Sequence 9, Appl
34	109.8	21.9	741	4 US-09-133-411-9	Sequence 9, Appl
35	109.8	21.9	2389	3 US-08-691-563C-52	Sequence 52, Appl
36	92.6	18.5	4480	4 US-09-167-322-12	Sequence 12, Appl
37	86.6	17.3	6363	2 US-08-929-967-6	Sequence 6, Appl
38	85.8	17.1	1140	4 US-09-603-185-5	Sequence 5, Appl
39	85.8	17.1	3612	4 US-09-265-013-3	Sequence 3, Appl
40	85.8	17.1	7308	4 US-09-011-745-3	Sequence 3, Appl
41	85.8	17.1	7308	4 US-09-011-745-4	Sequence 4, Appl
42	85.8	17.1	7616	4 US-09-011-745-2	Sequence 2, Appl
43	85.8	17.1	8202	1 US-08-258-420-13	Sequence 13, Appl
44	85.8	17.1	8332	3 US-08-850-961-1	Sequence 1, Appl
45	85.8	17.1	8332	4 US-09-479-776-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-471-724-1
Sequence 1, Application US/08471724
Patent No. 5800980
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Francois WALLEY
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BEDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV2 VIRUS AND MSRV2 PATHOGEN AND/OR
INJECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOLY
NUMBER OF INVENTION: 38
TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,724
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-471-724-1
Query Match: 83.8%; Score 419.6; DB 1; Length 1158;
Best Local Similarity: 91.9%; Pred. No. 2.5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;
QY 1 CAAGATCTCAGGATTCATGATGAGGCTGTGTTCTCTATAGCAGACTGTACTGACCT 60

Db 61 CAAGAACTCAGATATATCAATGAGGCTGTGTCTCTATACCAAGCTGTACTTAACCT 120
Qy 61 TATACCTGCTTCCCAATATCCAGAGAACAGAGGCTTTACAGCCCTGACCTTCAAG 120
Db 121 TATACAGGCTTTCCCAATATCCAGAGAACAGAGGCTTTACAGCTTGTGACCTTCAAG 180
Qy 121 GATGCTCTCTCTGATCCCTGATCATCTTCAATCTTGTGCTTGAAGAT 180
Db 181 GATGCTCTCTCTGATCCCTGATCATCTTCAATCTTGTGCTTGAAGAT 240
Qy 181 ACTTCAAAACCAATCTCACTCACTGACCTTAACTTCAACCAAGGCTTGAAGAT 240
Db 241 CCTTGAACCAACGCTCACTCACTGACCTTAACTTCAACCAAGGCTTGAAGAT 300
Qy 241 CCCCATCTATTTGGCAGGAGCTTGAACCAAGCTTGAACCAAGCTTGAAGAT 298
Db 301 CCCCATCTATTTGGCAGGAGCTTGAACCAAGCTTGAACCAAGCTTGAAGAT 360
Qy 299 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
Db 361 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 359 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGATGATGATGAT 418
Db 421 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGATGATGATGAT 480
Qy 419 AAGGCTCACTCTGCTACAGAG---GTTACTTGAAGGCTTAAATTTCCAAAGGAC 474
Db 481 AAGGCTCGGCTCTGCTACAGAGATTAGATGATGATGATGATGATGATGATGAT 540
Qy 475 AGGGCCCTCAGTGAAGAACATCCAG 501
Db 541 AGGGCCCTCAGTGAAGAACATCCAG 567

RESULT 2

US-08-471-969-1

Sequence 1, Application US/08471969

Patent No. 5871745

GENERAL INFORMATION:

APPLICANT: HEVE PERRON

APPLICANT: FRANCOIS MALLET

APPLICANT: BERNARD MANDRAND

APPLICANT: FREDERIC BEBINE

APPLICANT: FREDERIC BEBINE

TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL

TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,969

FILING DATE: June 6, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Berridge, William P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 36055A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1158 bases

TYPE: nucleotide

STRANDEDNESS: single-stranded

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-471-969-1

Query Match

83.8%; Score 419.6; DB 2; Length 1158;

Best Local Similarity 91.9%; Pred. No. 2.5e-140;

Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

Qy 1 CAAGATTCAGATATATCAATGAGGCTGTGTCTCTATACCAAGCTGTACTTAACCT 60
Db 61 CAAGAACTCAGATATATCAATGAGGCTGTGTCTCTATACCAAGCTGTACTTAACCT 120
Qy 61 TATACCTGCTTCCCAATATCCAGAGAACAGAGGCTTTACAGCCCTGACCTTCAAG 120
Db 121 TATACAGGCTTTCCCAATATCCAGAGAACAGAGGCTTTACAGCTTGTGACCTTCAAG 180
Qy 121 GATGCTCTCTCTGATCCCTGATCATCTTCAATCTTGTGCTTGAAGAT 180
Db 181 GATGCTCTCTCTGATCCCTGATCATCTTCAATCTTGTGCTTGAAGAT 240
Qy 181 ACTTCAAAACCAATCTCACTCACTGACCTTAACTTCAACCAAGGCTTGAAGAT 240
Db 241 CCTTGAACCAACGCTCACTCACTGACCTTAACTTCAACCAAGGCTTGAAGAT 300
Qy 241 CCCCATCTATTTGGCAGGAGCTTGAACCAAGCTTGAACCAAGCTTGAAGAT 298
Db 301 CCCCATCTATTTGGCAGGAGCTTGAACCAAGCTTGAACCAAGCTTGAAGAT 360
Qy 299 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
Db 361 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 359 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGATGATGATGAT 418
Db 421 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGATGATGATGAT 480
Qy 419 AAGGCTCACTCTGCTACAGAG---GTTACTTGAAGGCTTAAATTTCCAAAGGAC 474
Db 481 AAGGCTCGGCTCTGCTACAGAGATTAGATGATGATGATGATGATGATGATGAT 540
Qy 475 AGGGCCCTCAGTGAAGAACATCCAG 501
Db 541 AGGGCCCTCAGTGAAGAACATCCAG 567

RESULT 3

US-08-384-137-1

Sequence 1, Application US/08384137

Patent No. 5871996

GENERAL INFORMATION:

APPLICANT: HEVE PERRON

APPLICANT: FRANCOIS MALLET

APPLICANT: BERNARD MANDRAND

APPLICANT: FREDERIC BEBINE

APPLICANT: FREDERIC BEBINE

TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL

TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,137
FILING DATE: February 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-384-137-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140; Indels 6; Gaps 2;
Matches 466; Conservative 0; Mismatches 35;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATACCAAGCTGTACCTAGCCCT 60
DB 61 CAAGACTCAGGATTATCAATGAGGCTGTGTTCTCTATACCAAGCTGTACCTAGCCCT 120
QY 61 TATACCTGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 180
DB 181 GATGCTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 240
QY 181 ACTTCAAAACCAACATCTCACTCACTGATCTATTTTACCCCAAGGCTTACGAGATAGT 240
DB 241 CCTTGAACCCCAAGCTCTCACTCACTGATCTATTTTACCCCAAGGCTTACGAGATAGT 300
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGACTTGAAGCAATCTCTACCTGAGACA-- 298
DB 301 CCCCATCTATTTGGCCAGGATTAAGCCCAAGACTTGAAGCAATCTCTACCTGAGACT 360
QY 299 CTTGCTCTTGGTAGTGATGATTTACTTTGGCCGCGCCATTTCAAGAACTTTGGCCAT 358
DB 361 CTTGCTCTTGGTAGTGATGATTTACTTTGGCCGCGCCATTTCAAGAACTTTGGCCAT 420
QY 359 CAAGCACCCCAAGGCTCTTCAATTTCCGCTACCTGATGCTACATGTTTCCAAACCA 418
DB 421 CAAGCACCCCAAGGCTCTTCAATTTCCGCTACCTGATGCTACATGTTTCCAAACCA 480
QY 419 AAGGCTCAACTCTGCTCAACAGAG---GTTACTTAAAGGCTTAAATTTATCCAAAGGACC 474
DB 481 AAGGCTCGGCTCTGCTCAACAGAGATTAGATCTAAGGCTTAAATTTATCCAAAGGACC 540
QY 475 AAGGCTCTCACTGAGAGAACATCCAG 501
DB 541 AAGGCTCTCACTGAGAGAACATCCAG 567

RESULT 4
US-08-470-006A-1
Sequence 1, Application US/08470006A
GENERAL INFORMATION:
APPLICANT: Heive PERRON
APPLICANT: Brandois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BEDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND
NUMBER OF INVENTION: BIOPOLYMER CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,006A
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-470-006A-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140; Indels 6; Gaps 2;
Matches 466; Conservative 0; Mismatches 35;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATACCAAGCTGTACCTAGCCCT 60
DB 61 CAAGACTCAGGATTATCAATGAGGCTGTGTTCTCTATACCAAGCTGTACCTAGCCCT 120
QY 61 TATACCTGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 180
DB 181 GATGCTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 240
QY 181 ACTTCAAAACCAACATCTCACTCACTGATCTATTTTACCCCAAGGCTTACGAGATAGT 240
DB 241 CCTTGAACCCCAAGCTCTCACTCACTGATCTATTTTACCCCAAGGCTTACGAGATAGT 300
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGACTTGAAGCAATCTCTACCTGAGACA-- 298
DB 301 CCCCATCTATTTGGCCAGGATTAAGCCCAAGACTTGAAGCAATCTCTACCTGAGACT 360
QY 299 CTTGCTCTTGGTAGTGATGATTTACTTTGGCCGCGCCATTTCAAGAACTTTGGCCAT 358
DB 361 CTTGCTCTTGGTAGTGATGATTTACTTTGGCCGCGCCATTTCAAGAACTTTGGCCAT 420
QY 359 CAAGCACCCCAAGGCTCTTCAATTTCCGCTACCTGATGCTACATGTTTCCAAACCA 418
DB 421 CAAGCACCCCAAGGCTCTTCAATTTCCGCTACCTGATGCTACATGTTTCCAAACCA 480
QY 419 AAGGCTCAACTCTGCTCAACAGAG---GTTACTTAAAGGCTTAAATTTATCCAAAGGACC 474
DB 481 AAGGCTCGGCTCTGCTCAACAGAGATTAGATCTAAGGCTTAAATTTATCCAAAGGACC 540
QY 475 AAGGCTCTCACTGAGAGAACATCCAG 501
DB 541 AAGGCTCTCACTGAGAGAACATCCAG 567

RESULT 5

US-08-691-563C-1

; Sequence 1, Application US/08691563C
; Patent No. 6001987

; GENERAL INFORMATION:

; APPLICANT: HERVE PERRON

; APPLICANT: FREDERIC BESEME

; APPLICANT: FREDERIC BEDIN

; APPLICANT: GLAUCIA PARANHOS-BACCALA

; APPLICANT: FLORENCE KOMURIAN-PRADEL

; APPLICANT: COLETTE JOLIVET

; APPLICANT: BERNARD MANDRAND

; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC

; TITLE OF INVENTION: THERAPEUTIC PURPOSES

; NUMBER OF SEQUENCES: 92

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OLIFF & BERRIDGE

; STREET: 700 South Washington Street, Suite 300

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22314

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/691,563C

; FILING DATE: 02-AUG-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Berridge, William P.

; REGISTRATION NUMBER: 30,024

; REFERENCE/DOCKET NUMBER: WPB 38588

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-836-6400

; TELEFAX: 703-836-2787

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1158 base pairs

; TYPE: nucleotide

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

; US-08-691-563C-1

; Query Match

; Best Local Similarity 91.9%; Score 419.6; DB 3; Length 1158;

; Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

; Db 1 CAAGATTCAGATTATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGTACTGACCTT 60

; Db 61 CAAGAACTCAGATATATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGTACTGACCTT 120

; Db 61 TATACCTGCTTCCCAATATCCAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

; Db 121 TATACAGTCTTCCCAATATCCAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180

; Db 121 GATGCTTCTTGCATCCCTGTACGCTGTACGCTGTACGCTGTACGCTGTACGCTGTACG 240

; Db 181 GATGCTTCTTGCATCCCTGTACGCTGTACGCTGTACGCTGTACGCTGTACGCTGTACG 240

; Db 181 ACTTCAACCCCAATCTCACTGAGCTATTTTACCCCAAGGTTTCAAGGATGAT 240

; Db 241 CTTTGAACCAAGCTCTCACTGAGCTATTTTACCCCAAGGTTTCAAGGATGAT 300

; Db 241 CCGCATCTATTTGGCCAGAGATTTAGCCCAATCTCTATACCTGAGCA-- 298

; Db 301 CCGCATCTATTTGGCCAGAGATTTAGCCCAATCTCTATACCTGAGCACT 360

; Db 299 CTGTGCTTGGAGTATGATGATTTACTTTGGCCGCAATTCAGAAACCTTGTGCAT 358

; Db 361 CTGTGCTTGGAGTATGATGATTTACTTTGGCCGCAATTCAGAAACCTTGTGCAT 420

; Db 359 CAAGCCACCCCAAGGCTTCTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418

; Db 421 CAAGCCACCCCAAGGCTTCTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

; Db 419 AAGGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 474

; Db 481 AAGGCTCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540

; Db 475 AAGGCTCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 501

; Db 541 AAGGCTCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 567

RESULT 6

US-09-200-990-1

; Sequence 1, Application US/09200990
; Patent No. 6184025

; GENERAL INFORMATION:

; APPLICANT: HERVE PERRON

; APPLICANT: FRANCOIS MALLET

; APPLICANT: BERNARD MANDRAND

; APPLICANT: FREDERIC BEDIN

; APPLICANT: FREDERIC BESEME

; TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

; TITLE OF INVENTION: INSPECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOLY

; TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)

; NUMBER OF SEQUENCES: 38

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OLIFF & BERRIDGE

; STREET: 700 South Washington Street, Suite 300

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22314

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/200,990

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/471,965

; FILING DATE: June 6, 1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Berridge, William P.

; REGISTRATION NUMBER: 30,024

; REFERENCE/DOCKET NUMBER: WPB 36055A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-836-6400

; TELEFAX: 703-836-2787

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1158 bases

; TYPE: nucleotide

; STRANDEDNESS: single-stranded

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

; US-09-200-990-1

; Query Match

; Best Local Similarity 91.9%; Score 419.6; DB 4; Length 1158;

; Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

; Db 1 CAAGATTCAGATTATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGTACTGACCTT 60

; Db 61 CAAGAACTCAGATATATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGTACTGACCTT 120

QY 61 TATAGCTGCTTTCCCAATACCAAGAGAGAGAGTGTATTACAGTCTGACCTTCAG 120
DB 121 TATAGAGTGTCTTCCCAATATACCAAGAGAGAGAGTGTATTACAGTCTGACCTTCAG 180
QY 121 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 181 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 181 ACTTCAAAACCAACATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCT 240
DB 241 CTTTGAACCCCAAGCTCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCT 300
QY 241 CCCCATCTATTGTCGACGATTAAGCCCAAGCTTGAAGCTTCAATCTGACCA-- 298
DB 301 CCCCATCTATTGTCGACGATTAAGCCCAAGCTTGAAGCTTCAATCTGACCACT 360
QY 299 CTTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 361 CTTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 359 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGAT 418
DB 421 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGAT 480
QY 419 AAGGCTCACTCTGCTCAACAGCAG---GTTACTTAGGGCTAAATATTCACCAAGGCA 474
DB 481 AAGGCTCGGCTCTGCTCAACAGGATGATGATGATGATGATGATGATGATGATGAT 540
QY 475 AAGGCTCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 501
DB 541 AAGGCTCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 567

RESULT 7

US-09-133-411-1
Sequence 1, Application US/09133411

Patent No. 6342383

GENERAL INFORMATION:

APPLICANT: Hervé PERRON

APPLICANT: Francois MALLER

APPLICANT: Bernard MANDRAND

APPLICANT: Frederic BESEME

APPLICANT: Florence KOMORIAN-PRADEL

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-133-411-1

Query Match 83.8%; Score 419.6; DB 4; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTTATCAATGAGGCTGTTGCTCTATAGCCAGCTGATCCTGACCT 60
DB 61 CAAGATCTCAGGATTTATCAATGAGGCTGTTGCTCTATAGCCAGCTGATCCTGACCT 120
QY 61 TATAGCTGCTTTCCCAATATACCAAGAGAGAGTGTATTACAGTCTGACCTTCAG 120
DB 121 TATAGAGTGTCTTCCCAATATACCAAGAGAGAGTGTATTACAGTCTGACCTTCAG 180
QY 121 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 181 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 181 ACTTCAAAACCAACATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCT 240
DB 241 CTTTGAACCCCAAGCTCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCT 300
QY 241 CCCCATCTATTGTCGACGATTAAGCCCAAGCTTGAAGCTTCAATCTGACCA-- 298
DB 301 CCCCATCTATTGTCGACGATTAAGCCCAAGCTTGAAGCTTCAATCTGACCACT 360
QY 299 CTTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 361 CTTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 359 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGAT 418
DB 421 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGAT 480
QY 419 AAGGCTCACTCTGCTCAACAGCAG---GTTACTTAGGGCTAAATATTCACCAAGGCA 474
DB 481 AAGGCTCGGCTCTGCTCAACAGGATGATGATGATGATGATGATGATGATGATGAT 540
QY 475 AAGGCTCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 501
DB 541 AAGGCTCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 567

RESULT 8

US-08-691-563C-57
Sequence 57, Application US/08691563C

Patent No. 6001987

GENERAL INFORMATION:

APPLICANT: Hervé PERRON

APPLICANT: Frederic BESEME

APPLICANT: Florence KOMORIAN-PRADEL

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

APPLICANT: Bernard MANDRAND

APPLICANT: Colette JOLIVET

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 2391 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-691-563C-57

Query Match 83.8%; Score 419.6; DB 3; Length 2391;
Best Local Similarity 91.9%; Pred. No. 3.8e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 646 CAGGAATCTCAGGATTATCAATGAGGCTGTTCTCTATAGCCAGCTGTACCTAGCCCT 705
QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 706 TATACAGTCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 121 GATGCTTTCTTCTGCAATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCT 180
DB 766 GATGCTTTCTTCTGCAATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCT 825
QY 181 ACTTAAACCAACATCTCAATCTGAGCTATTTTACCCCAAGGTTTCAGGAGTACT 240
DB 826 CTTTGAACCAACATCTCAATCTGAGCTATTTTACCCCAAGGTTTCAGGAGTACT 885
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGATTGAGCCAACTCTCATACCTGAGCA 298
DB 886 CCCCATCTATTTGGCCAGGATTTAGCCCAAGATTGAGCCAACTCTCATACCTGAGCA 945
QY 299 CTTTGTCTTGGTGTGATGATTTACTTTTGGCCGCCATTTACAGAACTTTTGGCCAT 358
DB 946 CTTTGTCTTGGTGTGATGATTTACTTTTGGCCGCCATTTACAGAACTTTTGGCCAT 1005
QY 359 CAGGACCAACGAGCTCTCAATTTCTGAGCTGTGAGCTGAGCTGAGCTGAGCTGAGCT 418
DB 1006 CAGGACCAACGAGCTCTCAATTTCTGAGCTGTGAGCTGAGCTGAGCTGAGCTGAGCT 1065
QY 419 AAGGCTCACTGTCTACAG 474
DB 1066 AAGGCTCACTGTCTACAG 1125
QY 475 AGGGCCCTCAGTGAAGACATCCAG 501
DB 1126 AGGGCCCTCAGTGAAGACATCCAG 1152

RESULT 9
US-08-691-563C-89
Sequence 89, Application US/08691563C
Patent No. 6001987
GENERAL INFORMATION:
APPLICANT: Hervé BERRON
APPLICANT: Frédéric BESSEME
APPLICANT: Frédéric BEDIN
APPLICANT: Gaucia PARANHOS-BACCALA

APPLICANT: Florence KOMRIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 1577 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-691-563C-89

Query Match 80.7%; Score 404.4; DB 3; Length 1577;
Best Local Similarity 89.1%; Pred. No. 8.3e-135;
Matches 449; Conservative 0; Mismatches 51; Indels 4; Gaps 1;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 998 CAGGATCTCAGGATTATCAATGAGGCTGTTCTCTATAGCCAGCTGTACCTAGCCCT 1057
QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 1058 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
QY 121 GATGCTTTCTTCTGCAATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCT 180
DB 1058 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
QY 121 GATGCTTTCTTCTGCAATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCT 180
DB 1118 GATGCTTTCTTCTGCAATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCT 1177
QY 181 ACTTAAACCAACATCTCAATCTGAGCTATTTTACCCCAAGGTTTCAGGAGTACT 240
DB 1178 CTTTGAACCAACATCTCAATCTGAGCTATTTTACCCCAAGGTTTCAGGAGTACT 1237
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGATTGAGCCAACTCTCATACCTGAGCA 300
DB 1238 CCCCATCTATTTGGCCAGGATTTAGCCCAAGATTGAGCCAACTCTCATACCTGAGCA 1297
QY 301 TGTCTTGTGAGTGTGATGATTTACTTTTGGCCGCCATTTACAGAACTTTTGGCCATCA 360
DB 1238 TGTCTTGTGAGTGTGATGATTTACTTTTGGCCGCCATTTACAGAACTTTTGGCCATCA 1357
QY 361 AGCCACCAAGGCTCTTCAATTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 1358 AGCCACCAAGGCTCTTCAATTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1417
QY 421 GGGCTCACTGTCTACAG 476
DB 1418 GGGCTCACTGTCTACAG 1477

QY 477 GGCCCTCAGTGAGAACATCCCA 500
Db 1478 GGCCCTCTGTGAGGAATGATCCCA 1501

RESULT 10

US-09-120-653D-4
; Sequence 4, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-MON
; APPLICANT: JON, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120, 653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 4
; LENGTH: 2330
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-4

Query Match 54.3%; Score 271.8; DB 4; Length 2330;
Best Local Similarity 73.8%; Pred. No. 3.7e-87;
Matches 374; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTTCTCTATACCCAGCTGTACCTGACCT 60
Db 624 CAAGATCTTAACTACATCAATGAGGCTGTTCTCTATACCCAGCTGTACCTGACCT 683
QY 61 TATCTCGCTTCCCAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 684 TATACCTGCTCTCTCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 743
QY 121 GATGCTTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 180
Db 744 GATGCTTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 803
QY 181 ACTTCAACCAACATCTCACTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 240
Db 804 CCCAACCAACCAACATCTCACTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 863
QY 241 CCCCATCTATTGCGCAGGATGAGCCCAAGATGAGCCCAAGATGAGCCCAAGATGAGCC 298
Db 864 CCTATCTGTTGGTCAAGGCTTCAAGATGAGCCCAAGATGAGCCCAAGATGAGCCCA 923
QY 299 CTGTGCTTGGTGGATGATGATTTACTTTGGCGCCCAATTGAGAAGCTTGGCCAT 358
Db 924 CTGTGCTTCAATATGATGATGATTTACTTTGGCTACAGATGAGAAGCTTGGCCAG 983
QY 359 CAAGCCACCAAGGCTTCAATTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCT 418
Db 984 CAGGCTACTAGATCTCTGGAACCTTCTAGCTAATCAAGGCTGCTGCTGCTGCTGCTGCT 1043
QY 419 AAGGCTCAACTGCTGCTACAGAGT----TACTTAGGGCTTAAATTTCCAAAGGCA 474
Db 1044 AAGGCCCAACTTGTCTACAGAGGTTAAATATCTTAGGCTTAAATTTCCAAAGGCA 1103
QY 475 AGGGCCCTCAGTGAGAACATCCAG 501
Db 1104 AGGGCCCTCAGTGAGAACATCCAG 1130

RESULT 11

US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-MON
; APPLICANT: JON, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120, 653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 1
; LENGTH: 3910
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

Query Match 54.3%; Score 271.8; DB 4; Length 3910;
Best Local Similarity 73.8%; Pred. No. 5e-87;
Matches 374; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTTCTCTATACCCAGCTGTACCTGACCT 60
Db 1020 CAAGATCTTAACTACATCAATGAGGCTGTTCTCTATACCCAGCTGTACCTGACCT 1079
QY 61 TATCTCGCTTCCCAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 1080 TATACCTGCTCTCTCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1139
QY 121 GATGCTTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 180
Db 1140 GATGCTTCTTCTGATCCTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 1199
QY 181 ACTTCAACCAACATCTCACTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 240
Db 1200 CCCAACCAACCAACATCTCACTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 1259
QY 241 CCCCATCTATTGCGCAGGATGAGCCCAAGATGAGCCCAAGATGAGCCCAAGATGAGCC 298
Db 1260 CCTATCTGTTGGTCAAGGCTTCAAGATGAGCCCAAGATGAGCCCAAGATGAGCCCA 1319
QY 299 CTGTGCTTGGTGGATGATGATTTACTTTGGCGCCCAATTGAGAAGCTTGGCCAT 358
Db 1320 CTGTGCTTCAATATGATGATGATTTACTTTGGCTACAGATGAGAAGCTTGGCCAG 1379
QY 359 CAAGCCACCAAGGCTTCAATTTCTGCTACCTGCTGCTGCTGCTGCTGCTGCTGCT 418
Db 1380 CAGGCTACTAGATCTCTGGAACCTTCTAGCTAATCAAGGCTGCTGCTGCTGCTGCTGCT 1439
QY 419 AAGGCTCAACTGCTGCTACAGAGT----TACTTAGGGCTTAAATTTCCAAAGGCA 474
Db 1440 AAGGCCCAACTTGTCTACAGAGGTTAAATATCTTAGGCTTAAATTTCCAAAGGCA 1499
QY 475 AGGGCCCTCAGTGAGAACATCCAG 501
Db 1500 AGGGCCCTCAGTGAGAACATCCAG 1526

RESULT 12

US-08-471-724-2
; Sequence 2, Application US/08471724
; Patent No. 580980
; GENERAL INFORMATION:
; APPLICANT: HEVE PERRON
; APPLICANT: FRANCOIS MALLEET
; APPLICANT: BERNARD MANDRAND

APPLICANT: Frederic BREDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,724
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-471-724-2

Query Match 42.4%; Score 212.6; DB 1; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGATTATCAATGAGGCTGTGTCTCTATAGCCAGCTGTACTAGCCCT 60
DB 61 CAAGAACTCAGATTATCAATGAGGCTGTGTCTCTATAGCCAGCTGTACTAGCCCT 120
QY 61 TATACCTGCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGTCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTTCTTGATGATCCCTGTACATCTGACCTCAATCTGTGTTCCTTTGAAGAT 180
DB 181 GATGCTTCTTGATGATCCCTGTACATCTGACCTCAATCTGTGTTCCTTTGAAGAT 240
QY 181 ACTTCAACCCCAAGATCTCACTGAGTACTTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAAGCTCACTGAGTACTTATTTTACCCCAAGGTTCAAGG 295

RESULT 13
US-08-471-969-2
Sequence 2, Application US/08471969
Patent No. 5871745
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Francois MALLLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BREDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:

ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,969
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-471-969-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGATTATCAATGAGGCTGTGTCTCTATAGCCAGCTGTACTAGCCCT 60
DB 61 CAAGAACTCAGATTATCAATGAGGCTGTGTCTCTATAGCCAGCTGTACTAGCCCT 120
QY 61 TATACCTGCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGTCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTTCTTGATGATCCCTGTACATCTGACCTCAATCTGTGTTCCTTTGAAGAT 180
DB 181 GATGCTTCTTGATGATCCCTGTACATCTGACCTCAATCTGTGTTCCTTTGAAGAT 240
QY 181 ACTTCAACCCCAAGATCTCACTGAGTACTTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAAGCTCACTGAGTACTTATTTTACCCCAAGGTTCAAGG 295

RESULT 14
US-08-384-137-2
Sequence 2, Application US/08384137
Patent No. 5871996
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Francois MALLLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BREDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,137
FILING DATE: February 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-384-137-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarly 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCGCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTAACCTAGCCCT 60
DB 61 CAAGACTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTAACCTAGCCCT 120
QY 61 TATACCTGCTTTCCCAATATCAGAGAAACAGAGTGTATACAGTCTGGAACCTTCAAG 120
DB 121 TATACAGTGTCTTCCCAATATCAGAGAAACAGAGTGTATACAGTCTGGAACCTTCAAG 180
QY 121 GATGCTTTCTTGCATCCCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCT 180
DB 181 GATGCTTTCTTGCATCCCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCT 240
QY 181 ACTTCAACCAACATCTCACTGACCTGACCTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAAGCTCACTGACCTGACCTGTTTACCCCAAGGTTCAAGG 295

RESULT 15

US-08-470-006A-2
Sequence 2, Application US/08470006A
Patent No. 5962217
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BEDIN
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INEFFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND
TITLE OF INVENTION: BIOPOLYMER CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,006A
FILING DATE: June 6, 1995

ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-470-006A-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarly 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCGCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTAACCTAGCCCT 60
DB 61 CAAGACTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTAACCTAGCCCT 120
QY 61 TATACCTGCTTTCCCAATATCAGAGAAACAGAGTGTATACAGTCTGGAACCTTCAAG 120
DB 121 TATACAGTGTCTTCCCAATATCAGAGAAACAGAGTGTATACAGTCTGGAACCTTCAAG 180
QY 121 GATGCTTTCTTGCATCCCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCT 180
DB 181 GATGCTTTCTTGCATCCCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCT 240
QY 181 ACTTCAACCAACATCTCACTGACCTGACCTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAAGCTCACTGACCTGACCTGTTTACCCCAAGGTTCAAGG 295

Search completed: April 19, 2003, 14:48:59
Job time : 38.6897 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 19, 2003, 12:08:18 ; Search time 75.717 Seconds
(without alignments)
6646.925 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500
Perfect score: 501
Sequence: 1 caagatctcagatcatca.....tcagtgcagacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCITS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	431.8	86.2	1600	7	US-08-979-847-207
2	425.6	85.0	1600	7	US-08-979-847-206
3	422.2	84.3	2304	7	US-08-979-847-87
4	419.6	83.8	1158	7	US-08-979-847-1
5	419.6	83.8	1597	7	US-08-979-847-205
6	419.6	83.8	2365	7	US-08-979-847-88
7	419.6	83.8	2391	7	US-08-979-847-53
8	404.4	80.7	1577	7	US-08-979-847-83
9	402.2	80.3	568	10	US-08-864-761-16537
10	236.6	47.2	557	10	US-08-864-761-11981
11	231.6	46.2	449	10	US-08-864-761-11921
12	221.6	44.2	297	10	US-08-864-761-1194
13	212.6	42.4	297	7	US-08-979-847-2
14	212	42.3	540	10	US-08-864-761-15266
15	211.4	42.2	645	7	US-08-979-847-8
16	182.6	36.4	438	7	US-08-979-847-139
17	182.6	36.4	438	7	US-08-979-847-130
18	182.2	36.4	438	7	US-08-979-847-128
19	182.2	36.4	438	7	US-08-979-847-131

20	181.8	36.3	438	7	US-08-979-847-201	Sequence 201, App
21	165.6	33.1	944	10	US-09-864-761-14911	Sequence 14911, A
22	146.6	29.3	429	7	US-08-979-847-138	Sequence 138, App
23	145.4	29.0	429	7	US-08-979-847-135	Sequence 135, App
24	145	28.9	429	7	US-08-979-847-136	Sequence 136, App
25	145	28.9	429	7	US-08-979-847-137	Sequence 137, App
26	137.2	27.4	150	7	US-08-979-847-36	Sequence 36, App1
27	116.4	23.2	2448	7	US-08-979-847-49	Sequence 49, App1
28	114.2	22.8	191	10	US-09-864-761-119702	Sequence 19702, A
29	109.8	21.9	741	7	US-08-979-847-9	Sequence 9, App1
30	109.8	21.9	2389	7	US-08-979-847-48	Sequence 48, App1
31	107.6	21.5	133	10	US-09-864-761-33921	Sequence 33021, A
32	104.4	20.8	431	10	US-09-864-761-28100	Sequence 28100, A
33	102.2	20.4	181	10	US-09-864-761-17958	Sequence 17958, A
34	92.6	18.5	447	10	US-09-864-761-11512	Sequence 11512, A
35	92.4	18.4	326014	10	US-09-731-231A-3	Sequence 3, App1
36	89	17.8	157	10	US-09-864-761-31788	Sequence 31788, A
37	85.8	17.1	2151	10	US-09-845-157-1	Sequence 1, App1
38	85.8	17.1	8332	10	US-09-006-298-1	Sequence 1, App1
39	82.8	16.5	143068	10	US-09-967-768A-316	Sequence 316, App1
40	77.8	15.5	8323	10	US-09-970-597-2	Sequence 2, App1
41	76.8	15.3	183	10	US-09-864-761-28561	Sequence 28561, A
42	76	15.2	685	10	US-09-864-761-31430	Sequence 31430, A
43	70.8	14.1	1436	10	US-09-864-761-17414	Sequence 17414, A
44	69.8	13.9	85	7	US-08-979-847-195	Sequence 195, App
45	69.4	13.9	593	10	US-09-864-761-7065	Sequence 7065, App

ALIGNMENTS

RESULT 1
US-08-979-847-207
Sequence 207, Application US/08979847
Publication NO. US20030039664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESIME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMRIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER-READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979, 847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 207:

SEQUENCE CHARACTERISTICS:
LENGTH: 1600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-979-847-207

Query Match 86.2%; Score 431.8; DB 7; Length 1600;
Best Local Similarity 90.5%; Pred. No. 5,5e-133;
Matches 459; Conservative 20; Mismatches 22; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGATTATCATGAGGCTGTGTCTCTATATACCAAGCTGTACCTTACCTT 60
DB 373 CAAGAACTCAGATTATCATGAGGCTGTGTCTCTATATACCAAGCTGTACCTTACCTT 432
QY 61 TATATCTGCTTCTCCCAATATACCAAGAAAGTGAAGTGTATACAGTCTGTGACCTTACG 120
DB 433 TATACAGTCTTCTCCCAATATACCAAGAAAGTGAAGTGTATACAGTCTGTGACCTTACG 492
QY 121 GATGCTTCTCTGATCCCTGTACATCTGATCTCAATCTGTGTGCTTGTGACAT 180
DB 493 GATGCTTCTCTGATCCCTGTACATCTGATCTCAATCTGTGTGCTTGTGACAT 552
QY 181 ACTTCAAAACCAACATCTCACTCACTGACTATTTTACCCCAAGGTTCAAGGATAGT 240
DB 553 CTTTGAACCCCAACGCTCACTCACTGACTATTTTACCCCAAGGTTCAAGGATAGT 612
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCAATCTCTATCTGTGACA-- 298
DB 613 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCAATCTCTATCTGTGACA-- 672
QY 299 CTTGCTCTCGGTAGGTGATGATTACTTTTGKCCGCCCATTCAGAACTTTGTGCCAT 358
DB 673 CTTGCTCTCGGTAGGTGATGATTACTTTTGKCCGCCCATTCAGAACTTTGTGCCAT 732
QY 359 CAAGCAACCAAGCGCTCTCAATTTCTCTGCTACCTGTGGCTACATGTTTCCAAACCA 418
DB 733 CAAGCAACCAAGCGCTCTCAATTTCTCTGCTACCTGTGGCTACATGTTTCCAAACCA 792
QY 419 AAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTTAAATTTCCAAAGGCACC 474
DB 793 AAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTTAAATTTCCAAAGGCACC 852
QY 475 AGGGCCCTCAGTGAAGAACATCCAG 501
DB 853 AGGGCCCTCAGTGAAGAACATCCAG 879

RESULT 2

US-08-979-847-206
; Sequence 206, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESS: OLIVIER & BERRIDGE, PLC
; STREET: P. O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA

ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 206:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-979-847-206

Query Match 85.0%; Score 425.6; DB 7; Length 1600;
Best Local Similarity 88.6%; Pred. No. 6,3e-131;
Matches 449; Conservative 26; Mismatches 24; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGATTATCATGAGGCTGTGTCTCTATATACCAAGCTGTACCTTACCTT 60
DB 373 CAAGAACTCAGATTATCATGAGGCTGTGTCTCTATATACCAAGCTGTACCTTACCTT 432
QY 61 TATATCTGCTTCTCCCAATATACCAAGAAAGTGAAGTGTATACAGTCTGTGACCTTACG 120
DB 433 TATACAGTCTTCTCCCAATATACCAAGAAAGTGAAGTGTATACAGTCTGTGACCTTACG 492
QY 121 GATGCTTCTCTGATCCCTGTACATCTGATCTCAATCTGTGTGCTTGTGACAT 180
DB 493 GATGCTTCTCTGATCCCTGTACATCTGATCTCAATCTGTGTGCTTGTGACAT 552
QY 181 ACTTCAAAACCAACATCTCACTCACTGACTATTTTACCCCAAGGTTCAAGGATAGT 240
DB 553 CTTTGAACCCCAACGCTCACTCACTGACTATTTTACCCCAAGGTTCAAGGATAGT 612
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCAATCTCTATCTGTGACA-- 298
DB 613 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCAATCTCTATCTGTGACA-- 672
QY 299 CTTGCTCTCGGTAGGTGATGATTACTTTTGKCCGCCCATTCAGAACTTTGTGCCAT 358
DB 673 CTTGCTCTCGGTAGGTGATGATTACTTTTGKCCGCCCATTCAGAACTTTGTGCCAT 732
QY 359 CAAGCAACCAAGCGCTCTCAATTTCTCTGCTACCTGTGGCTACATGTTTCCAAACCA 418
DB 733 CAAGCAACCAAGCGCTCTCAATTTCTCTGCTACCTGTGGCTACATGTTTCCAAACCA 792
QY 419 AAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTTAAATTTCCAAAGGCACC 474
DB 793 AAGGCTCACTCTGCTCAACAGAGT---TACTTAGGGCTTAAATTTCCAAAGGCACC 852
QY 475 AGGGCCCTCAGTGAAGAACATCCAG 501
DB 853 AGGGCCCTCAGTGAAGAACATCCAG 879

RESULT 3

US-08-979-847-87
; Sequence 87, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE

APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 87:
SEQUENCE CHARACTERISTICS:
LENGTH: 2304 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
US-08-979-847-87

Query Match 84.3%; Score 422.2; DB 7; Length 2304;
Best Local Similarity 92.3%; Pred. No. 9.9e-130;
Matches 468; Conservative 0; Mismatches 33; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 640 CAAGATCTCAGGATTATTAATGAGGCTGTTTCTCTATATACCCAGCTGTATCTAGCCCT 699
QY 61 TATATCTGCTTTCCCAATATCCAGAGAGAGAGGTGTTTACAGTCTGTGACCTTCAAG 120
DB 700 TATATCTGCTTTCCCTAATATCCAGAGAGAGAGGTGTTTACAGTCTGTGACCTTCAAG 759
QY 121 GATGCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 760 GATGCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 819
QY 181 ACTTCAAAACCAACATCTCAACTGAGTATTTTACCCCAAGGTTTCAGGGATAGT 240
DB 820 CTTTGAACCCCAAGCTCACTGAGTATTTTACCCCAAGGTTTCAGGGATAGT 879
QY 241 CCCCATCTATTTGGGAGGATTTAGCCCAAGCTGAGCCATCTCTCATCTGTGACCA-- 298
DB 880 CCCCATCTATTTGGGAGGATTTAGCCCAAGCTGAGCCATCTCTCATCTGTGACCACT 939
QY 299 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 940 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 999
QY 359 CAAGCACCACCAAGCGCTTCAATTTCTGCTGATCTGTGCTGATGATGTTTCAACCA 418

DB 1000 CAAGCACCACCAAGACTTCTTACTTCTCTACTGCTGTGCTCAAGGTTTCCAAACCA 1059
QY 419 AAGGCTCAACTGCTGCTACAGCAG-----GTTACTTAGGGCTAAATTTATCCAAAGCAC 474
DB 1060 AAGGCTGCGCTCTGCTCTACAGGATTTAGATCTTAGGGCTTAAATTTATCCAAAGCAC 1119
QY 475 AAGGCTCTACTGAGGAACACATCCAG 501
DB 1120 AAGGCTCTACTGAGGAACATCCAG 1146

RESULT 4
US-08-979-847-1
Sequence 1, Application US/08979847
Publication No. US20030039664A1
GENERAL INFORMATION:
APPLICANT: PERON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-1

Query Match 83.8%; Score 419.6; DB 7; Length 1158;
Best Local Similarity 91.9%; Pred. No. 5.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATATCTGCTTTCCCAATATCCAGAGAGAGAGGTGTTTACAGTCTGTGACCTTCAAG 120
DB 121 TATATGCTTTCCCAATATCCAGAGAGAGAGGTGTTTACAGTCTGTGACCTTCAAG 180

Qy	121	GAAGCCTCTTGATATCCCGTACATCTGACCTCTCAATCTTGTTGGCTTTGAAGAT	180
Db	181	GATGCTTTTCTTGATCCCTGAGCTCCTCACTCAATCTTGTTGGCTTTGAAGAT	240
Qy	181	ACTTCAAACCCAACTCACTCACTACCTGACTATTTTACCCCAAGGTTCAAGGATAGT	240
Db	241	CCTTTGAACCCAAAGTCTCACTCACTGAGCTATTATACCCCAAGGTTCAAGGATAGC	300
Qy	241	CCCCATCTATTTGGCCAGGCAATTAGCCAAAGCTTTAGACCATCTCTCAATCTGAGACA--	298
Db	301	CCCCATCTATTTGGCCAGGCAATTAGCCAAAGCTTTAGATCTATCTCATCTGAGACACT	360
Qy	299	CTTGCCCTTGCGTAGTGAATTAATTCTTTGACCGCCATTAGAAACCTTGTCAT	358
Db	361	CTTGCTCTCAATAGATGAATTAATTCTTTTATGCGCCGTTAGAAACCTTGTCAT	420
Qy	359	CAAGCACCACCAAGCGCTTTCATATTTCTCGCTACCTGTGCTACATGTTTCCAAACA	418
Db	421	CAAGCACCACCAAGAACTTTAACTTTCTCTACTCTGCTACAAAGTTTCCAAACA	480
Qy	419	AAGGCTCAACTCTGCTCAGACAG---GTTACTTAGGGCTAAATAATTCAAAGGCACC	474
Db	481	AAGGCTCGCTCTGCTCACAAGATTTAGATACINAGGGCTAAATAATTCCAAAGGCACC	540
Qy	475	AGGGCCCTCAGTGAAGAACATATCCAG	501
Db	541	AGGGCCCTCAGTGAAGAACATATCCAG	567

RESULT 5

US-08-979-847-205
; Sequence 205, Application US/08979847
; Publication No. US20030039664A1

GENERAL INFORMATION:

APPLICANT: PERON, HERVE
 APPLICANT: BESEME, FREDERIC
 APPLICANT: BEDIN, FREDERIC
 APPLICANT: PARANHOS-BACCALA, GLAUCIA
 APPLICANT: KOMURIAN-PRADEL, FLORENCE
 APPLICANT: JOLIVET-REYNAUD, COLETTE
 APPLICANT: MANDRAND, BERNARD
 APPLICANT: GARSON, JEREMY
 APPLICANT: TUKE, PHILIP
 TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
 TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
 TITLE OF INVENTION: THERAPEUTIC PURPOSES
 NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928

;
;
CITY: ALEXANDRIA
STATE: VA

COUNTRY: USA

```

;      ZIP:  22320
;      COMPUTER READABLE FORM:
;

```

MEDIUM TYPE: Floppy disk

```
;
;  COMPUTER:  IBM PC compatible
;  OPERATING SYSTEM:  PC-DOS/MS-DOS
;
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30.024

TELECOMMUNICATION INFORMATION:
; REFERENCE/DOCKET NUMBER: WPB 39046A
; NOTIFICATION NUMBER: 30,024

TELEPHONE: 703-836-6400
TELEFAX: 703-836-3787

TELEFAX: 703-836-2187
; INFORMATION FOR SEQ ID NO: 205:
SEQUENCE CHARACTERISTICS:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1597 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-979-847-205

```

Query Match	83.8%	Score 419.6;	DB 7;	Length 1597;
Best Local Similarity	91.9%	Pred. No. 6.1e-129;		
Matches 466; Conservative	0;	Mismatches 35;	Indels 6;	Gaps 2

Oy	1	CAAGATCTCAGGATTATCAATGAGGCGTTGTTCTCTATAGCCAGCTGAACCTAG	60
Db	373	CAAGAAGCTCAGAGATTATCAATGAGGCGTTGTTCTCTATAGCCAGCTGAACCT	432
Oy	61	TATACTCTGCTTTTCCCAATTCGAGAGAACAGAGTGTTTACAGTCTGGAACCTCAG	120
Db	433	TATACAGTCTTTTCCCAATTCGAGAGAACAGAGTGTTTACAGTCTGGAACCTTAC	492
Oy	121	GATGCGTCTCTTCGATCCCGTGCATCCGACCTCAATTTCTTGTTGGCTTGAAGAT	180
Db	493	GATGCGTCTTTCTCGATCCCGTGCATCCGACCTCAATTTCTTGTTGGCTTGAAGAT	552
Oy	181	ACTTCAAAACCAACATCTCAACTACCTGAGCTATTTTACCCCAAGGGTTCAAGGATAGT	240
Db	553	CCTTTGAACCCCAAGCTCTCAACTACCTGAGCTATTTTACCCCAAGGGTTCAAGGATAGC	612
Oy	241	CCCCATCTATTTGGCCAGGCACTTAGCCCAAGCTTGAGCCAATCTCATACCTGAGACA--	298
Db	613	CCCCATCTATTTGGCCAGGCACTTAGCCCAAGCTTGAGCCAATCTCATACCTGAGACACT	672
Oy	299	CTTGTCCTTGGTAGTGGAGATTTTACTTTTGGCGGCCCAATCAGAAACCTGTGTCAT	358
Db	673	CTTGTCCTTGGTAGTGGAGATTTTACTTTTGGCGGCCCAATCAGAAACCTGTGTCAT	732
Oy	359	CAAGGCAACCCCAAGGCGTCTTCAATTTTCTGCGTACCTGTGGCTACATGTTTCCAAACCA	418
Db	733	CAAGGCAACCCCAAGGCGTCTTCAATTTTCTGCGTACCTGTGGCTACATGTTTCCAAACCA	792
Oy	419	AAGGCTCAACTCTGCTCAGACAGCAG---GTTACTTAGGGCTAAATTTATCCAAAGGCACC	474
Db	793	AAGGCTGGGCTCTGCTCAGACAGATTAGATACTNAGGGCTAAATTTATCCAAAGGCACC	852
Oy	475	AGGGGCTCTCAGTAGAGAACACATCCAG	501
Db	853	AGGGGCTCTCAGTAGAGAACATCTATCCAG	879

RESULT 6

US-08-979-847-88
; Sequence 88, Application US/08979847

; Publication No. US20030039664A1
; GENERAL INFORMATION:

APPLICANT: PERRON ;

;
;
; APPLICANT: BESEME, FREDER
; APPLICANT: BEDIN, FREDER

APPLICANT: PARANHOS-BACCALA

APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE

APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY

APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:

ADDRESS: OLIFF & BERT
STREET. P O BOX 19928

STATE: VA
CITY: ALEXANDRIA
STREET: F.O. BOYD

STATE: VA
COUNTRY: USA
STD: 00000

```

;      ZIP:  22320
;      COMPUTER READABLE FORM:

```


[illegible]

RESULT 8

```

US-08-979-847-93
: Sequence 83, Application US/08979847
: Publication No. US2003003964A1
:
: GENERAL INFORMATION:
:
: APPLICANT: PERRON, HERVE
:
: APPLICANT: BESEME, FREDERIC
:
: APPLICANT: BEDIN, FREDERIC
:
: APPLICANT: PARANHOS-BACCALA, GLAUCIA
:
: APPLICANT: KOMURIAN-PRADEL, FLORENCE
:
: APPLICANT: JOLIVET-REYNAUD, COLETTE
:
: APPLICANT: MANDRAND, BERNARD
:
: APPLICANT: GARSON, JEREMY
:
: APPLICANT: TUKE, PHILIP
:
: TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
:
: TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
:
: NUMBER OF SEQUENCES: 210
:
: CORRESPONDENCE ADDRESS:
:

```

ADDRESSSEE: OLIFF & BERRIDGE, PLC
 STREET: P.O. BOX 19928
 CITY: ALEXANDRIA
 STATE: VA
 COUNTRY: USA
 ZIP: 22320
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/979,847
 FILING DATE: 26-NOV-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: BERRIDGE, WILLIAM P.
 REGISTRATION NUMBER: 30,024
 REFERENCE/DOCKET NUMBER: WPB 39046A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-836-6400
 TELEFAX: 703-836-2787
 INFORMATION FOR SEQ ID NO: 83:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1577 base pairs
 TYPE: nucleotide
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 US-08-979-847-83

Query Match	Similarity	Score	DB %	Length
Best Local	89.1%	Pred. No. 6	7e-124	
Matches	449	Conservative	0	Mismatches 51; Indels 4; Gaps 1
QY	1	CAAGATCTCAGATTTATCAATGAGGCTGTGTTGCCCTATAGCAGCTGTACCTAGCCCT	60	
DB	998	CAAGATCTCAGATTTATTTATGAGGTGTTTTCCTATACCGAGCTGTACTAGCCCT	1057	
QY	61	TATACTCTCTCTTCCCAATATACCAGGAGAGAGATGTTTTACGTCCTGAGCCTTCAAG	120	
DB	1058	TATACTCTCTCTTCCCTAATATACCAGGAGAGAGATTAATTTACACTCTGAGCCTTAAG	1117	
QY	121	GATGCTTCTTCTGCATCCCTGTACATCTGACTCCAAATCTGTGTTGCCCTTTGAAGAT	180	

Db	1118	GATCCCTCTTTGCAACCCCTGACCTCTGATCTCAATCTTGTTGCTTTGAGAGAT	1177
Qy	181	ACTTCAAACCCCAACATCTCACTCACTACCTGATATTTTAAACCCAGGGTTCAAGGATAGT	240
Db	1178	CCTTTGAACCCCAATGCTCTCAATTCACCTGGACCTGTTTAAACCCAGGGGTTCCGGGATAGC	1237
Qy	241	CCCCATCTATTTTGGCCAGGATTCAGCCCAAGACTTGAGCCATCTCTCATCTGAGCACT	300
Db	1238	CCCCATCTATTTTGGCCAGGATTCAGCCCAAGACTTGAGCCATCTCTCATCTGAGCACT	1237
Qy	301	TGTCCTTCGAGTGGATGATATTTACTTTTGGCCGCCATTGAGAAACCTTGTGCATCA	360
Db	1298	TGTGCTTCGGTATGGGATATTTATTTTATGACACCCGTTGAGAAACCTTGTGCATCA	1357
Qy	361	AGCCACCCAAAGCGCTTTCATTTTCTCGCTACCTGTGGCTACATGTTTCCAAACCAA	420
Db	1358	AGCCACCCAAAGCGTTTAAATTTCTCTCACTCGTGTGGCTACAAAGTTTCCAAACCAA	1417
Qy	421	GGCTCACTCTGCTCACAGAGGT-----ACTTAAAGGCTAAATATATCAAAAGGACAGG	476
Db	1418	GGCTCACTCTGCTCACAGAGGTAAATATCTTAAAGGTTTAAATATATCAAAAGGACAGG	1477
Qy	477	GGCCCTCAGTGAAGAAACAATCA 500	
Db	1478	GGCCCTCTGTGAAGAAATGATCA 1501	

RESULT 9

```

US-09-864-761-16537
: Sequence 16537, Application US/09864761
: Patent No. US20020048763A1
:
GENERAL INFORMATION:
:
APPLICANT: Penn, Sharron G.
:
APPLICANT: Rank, David R.
:
APPLICANT: Hanzel, David K.
:
APPLICANT: Chen, Wensheng
:
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
:
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
:
FILE REFERENCE: Aeomica-X-1
:
CURRENT APPLICATION NUMBER: US/09/864,761
:
CURRENT FILING DATE: 2001-05-23
:
PRIOR APPLICATION NUMBER: US 60/180,312
:
PRIOR FILING DATE: 2000-02-04
:
PRIOR APPLICATION NUMBER: US 60/207,456
:
PRIOR FILING DATE: 2000-05-26
:
PRIOR APPLICATION NUMBER: US 09/632,366
:
PRIOR FILING DATE: 2000-08-03
:
PRIOR APPLICATION NUMBER: GB 24263.6
:
PRIOR FILING DATE: 2000-10-04
:
PRIOR APPLICATION NUMBER: US 60/236,359
:
PRIOR FILING DATE: 2000-09-27
:
PRIOR APPLICATION NUMBER: PCT/US01/00666
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: PCT/US01/00667
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: PCT/US01/00664
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: PCT/US01/00669
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: PCT/US01/00665
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: PCT/US01/00668
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: PCT/US01/00663
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: PCT/US01/00662
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: PCT/US01/00661
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: PCT/US01/00670
:
PRIOR FILING DATE: 2001-01-30
:
PRIOR APPLICATION NUMBER: US 60/234,687
:
PRIOR FILING DATE: 2000-09-21
:

```

;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 16537
;; LENGTH: 568
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC007567.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
US-09-864-761-16537

Query Match 80.3%; Score 402.2; DB 10; Length 568;
Best Local Similarity 91.1%; Pred. No. 2.3e-123;
Matches 462; Conservative 0; Mismatches 38; Indels 7; Gaps 3;

Qy 1 CAAGATCTCAGATTATCATAGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
Db 58 CAAGATCTCAGATTATCATAGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 117
Qy 61 TATACCTGCTTTCCCAATACCGAGAGACAGAGTGCTTTACAGTCTGACCTTCAG 120
Db 118 TATACCTGCTTTCCCAATACCGAGAGACAGAGTGCTTTATATCTGACCTTCAG 177
Qy 121 GATGCTCTCTGATGATCCGTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 180
Db 178 GATGCTCTCTGATGATCCGTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 237
Qy 181 ACTTCAACCAACATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGATCTG 240
Db 238 CCTTGAACCAACATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGATCTG 297
Qy 241 CCCCATCTATTTGGCAGGATTTAGCCCAAGACTTGAACCAATCTCTACCTGAGCA-- 298
Db 298 CCCCATCTATTTGGCAGGATTTAGCCCAAGACTTGAACCAATCTCTACCTGAGCACT 357
Qy 299 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 358
Db 358 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 417
Qy 359 CAAGCACCACCAAGCGCTCTTCATTTCTGCTACCTGCTGATCATGATGATGATGATG 418
Db 418 CAAGCACCACCAAGCGCTCTTCATTTCTGCTACCTGCTGATCATGATGATGATGATG 477
Qy 419 AAGGCTCACTCTGCTCAAGAGG---TTACTTAGGGCTTAAATTTCCAAAGGACAC 474
Db 478 AA-GCTCAACTCTGCTCAAGAGGCTTAAATTTCCAAAGGACAC 536
Qy 475 AGGGCCCTCATGAGAGACATCCAG 501
Db 537 AGGGCCCTCATGAGAGACATCCAG 563

RESULT 10

US-09-864-761-11981
; Sequence 11981, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aesomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456

;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 11981
;; LENGTH: 557
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC007923.2
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.7
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.73
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-11981

Query Match 47.2%; Score 236.6; DB 10; Length 557;
Best Local Similarity 72.2%; Pred. No. 2e-68;
Matches 337; Conservative 0; Mismatches 124; Indels 6; Gaps 2;

Qy 41 AGCAGCTGTAAGCTTACCTTATCTGCTTTCCAAATACGAGAGACAGAGTGT 100
Db 1 ACCCAGTTATACCAACCCCTGTACCTGCTCTCTCAAAATACGAGAGAGAGATAT 60
Qy 101 TTACAGTCTGAGACTTCAGATGCTTTCTTGTGATGCCGTGATCTTACTTCAAT 160
Db 61 TCAGTGTCTAGACCTCAAGATGCTTCTTATATCCCTCAGCTGATCCAGT 120
Qy 161 TCTGTTGACCTTTGAGATGATCTTCAACCAACCAATCTCACTGACCTGACTATTTTAC 220
Db 121 TTCTCTTGTGCTTTGAGATGATCCACAGACACACATCCCAATCTTATGAGACAGCTTGC 180
Qy 221 CCAGAGGTTGAGGATAGTCCCATCTATTTTGGCCAGGATTTAGCCCAAGACTTGAGCC 280
Db 181 CCAGAGGTTTAGGATAGCCCTCATGTTTGTGACAGGCGCATGAGCCCAAGATCTAGACC 240
Qy 281 AATCTCA--TACCTGACACTTGTCTTGGTGAAGTGAATTTACTTTTGGCCGCC 338
Db 241 ACTTCTAAGTCAGGACACTGTGTTCTTCAAGTATGAGATGATTTACTTTTGGCTACCA 300

QY 339 ATTCAGAACTTGTGTCATCAAGCCACCAACGCTCTTCAATTCTCGTACTGTG 398
 Db 301 GTTGAAGACCTCATGTGAGCAGGCTACTCTAKATCTCTTGAATCTTAGCTATCAAG 360
 QY 399 GCTCATGTGTTCCAAACCAAGGCTCAACTCTGCTCAGAGAGT---TACTTAGGGC 454
 Db 361 GGTAAAGGACATCTAGGTGAGAGGCTCAGCTTACTCTACAGAGGCTCAAAATATCTAGGCC 420
 QY 455 TAAATATCCAAAGCAGCAGGGGCTCAGTGAAGAACATCCAG 501
 Db 421 TAACTTATCCAGAGGAGCAGGGGCTCTACAGAGGAATGATACAG 467
 RESULT 11
 US-09-864-761-2921/c
 ; Sequence 2921, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecomica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; PRIOR FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 2921
 ; LENGTH: 449
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC010856.2
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.6
 ; OTHER INFORMATION: EXPRESSED IN B7474, SIGNAL = 2.7
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
 ; US-09-864-761-2921
 Query Match 46.2%; Score 231.6; DB 10; Length 449;
 Best Local Similarity 72.8%; Pred. No. 8,4e-67;
 Matches 326; Conservative 0; Mismatches 119; Indels 3; Gaps 2;
 QY 41 AGCCAGCTGTACTAGCCCTTATCTCTGCTTTCCAAATACAGAGAGCAGAGTGT 100
 Db 449 ATCCAGGTATACCAACCCCTATACCTGCTCTCAAAATACAGAGAGCAGAGTGT 390
 QY 101 TTACAGTCTGACCTTCAGAGATGCTTTCTTGATCCCTGATCCTGACTCTCAAT 160
 Db 389 TCATTGTTCTGACCTCAAGATGCTTCTCTGATCTCTGACTCTGACCTCAAGG 330
 QY 161 TCTGTTGCTTGAAGATATCTTCAACCAATCTCAACTCACTGACCTGACTTTTAC 220
 Db 329 TTCTCTTGTCTTGAAGATCTCCACAGACACAGCTCCCACTTCTGATGCTTTGC 270
 QY 221 CCNAGGTTCCAGGATATGTCCTATTTTGGCAGGACTTAGCCCAAGCTTGAAGC 280
 Db 269 CTCAAGGTTTAGGGATATGCTTATCTGTTGTCAGGACCTGCCCAAGATCTAGGCC 210
 QY 281 AATCTCTA--TACTGACACTTGTCTTGGTAGTGATGATTTACTTTGGCCGCC 338
 Db 209 ACTTCTCAAGTACAGGACCTGTGCTTCAAGTAGGATTTACTTATGTTTACCA 150
 QY 339 ATTCAGAACTTGTGTCATCAAGCCACCAAGGCTCTTCAATTCTCGTACTGTG 398
 Db 149 GTTCAAAAGCTTCATGCGACAGAGCTACTCTAGCTCTCTTGAATCTTAGCTATCAAG 90
 QY 399 GCTCATGTGTTCCAAACCAAGGCTCAACTCTGCTCAC-AGCAGGTTACTTAGGGCTAA 457
 Db 89 GGTACATGGCATCTAAACTGAAGGCCAGGCTGCTCAAGTCAAAATATCTAGGCCCTAA 30
 QY 458 AATATCCAAAGCAGCAGGGGCTCTCAG 485
 Db 29 TCTTAGCCAGAGAACCAAGGCTCTCAG 2
 RESULT 12
 US-09-864-761-1194/c
 ; Sequence 1194, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecomica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; PRIOR FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 1194
 LENGTH: 467
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC006999.2
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.3
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.9
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.82
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
 US-09-864-761-1194
 Query Match 44.2%; Score 221.6; DB 10; Length 467;
 Best Local Similarity 70.7%; Pred. No. 1.8e-63; Indels 6; Gaps 2;
 Matches 325; Conservative 0; Mismatches 129; Indels 6; Gaps 2;
 48 TGTACTAGAGCCCTTATCTCTGCTTCCCAATACAGAGGAGAGAGTGTTCACAGT 107
 467 TGACCCCAACACCTATCCCTGCTCTCAATACAGAGGAGATCAGATGTTTCATGCT 408
 108 CTTGACCTTCAAGATGCTTCTTCTGTCATCCCTGATCCTGATCTCAATCTTCTT 167
 407 TCTGACCTCAAGATTAATCTTCTTATTCCTCTGACCTGATCTCCCAATTTCTCT 348
 168 TGCCTTGAAGATCTCAATCCCAACATCTCACTCACTCCGAGCTATTTAACCAGG 227
 347 TGCCTTGAAGATCCCAAGACACATGTCCCAATTAAGTGAGAGCTTGCCCAAGG 288
 228 GTTCAGGATAGTCCCATCTATTGTGGCCAGCATTTAGCCCAAGCTTGAGCCATCTC 287
 287 GTTAAAGGATAGGCTTCACTGTTTGTGCAAGCACTGACCCCAAGATCTAGCCATCTC 228
 288 ATACCTGG--AACTTGTCTTGGAGTGTGATGATTTACTTTGGCGGCCCATTCAGA 345
 227 AAATCCAGCCACTGTGCTTCAAGTATGTGATGATTTACTTTGGCTACAGTTCTCA 168
 346 AACCTTGCATCAAGCCCAAGGAGGCTTCAATTTCTGCTACTGCTGAGTCAAT 405
 167 AGCTCATGCAAGAGGCTACTCTAGATCTTTGAATCTTCTAGTAAATCAAGGGTAA 108
 406 GGTTCACCAACCAAGGCTCAACTCTGCTCAAGCAGG---TTACTTAGGGCTAAAT 461

Db 107 GCTGTTTAGTCAAAAGCCCGAGCTTTCCTACAGAGGCCAAATATCTAGCTTATCAT 48
 Qy 462 ATCAAGGACCAAGGCGCCCTCAGTGAGGAACATCAAC 501
 Db 47 AGCCAGAGGACCAAGGCGCTCAGCAAGGATGAAAACG 8
 RESULT 13
 US-08-979-847-2
 Sequence 2, Application US/08979847
 Publication No. US2003039664A1
 GENERAL INFORMATION:
 APPLICANT: PERON, HERVE
 APPLICANT: BESEME, FREDERIC
 APPLICANT: BEDIN, FREDERIC
 APPLICANT: PARANHOS-BACCALA, GLAUCIA
 APPLICANT: KOMURIAN-PRADEL, FLORENCE
 APPLICANT: JOLIVIER-REYNARD, COLETTE
 APPLICANT: MANDRAND, BERNARD
 APPLICANT: GARSON, JEREMY
 APPLICANT: TURE, PHILIP
 TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
 TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
 TITLE OF INVENTION: THERAPEUTIC PURPOSES
 NUMBER OF SEQUENCES: 210
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OLIVIER & BERRIDGE, PLC
 STREET: P.O. BOX 19928
 CITY: ALEXANDRIA
 STATE: VA
 COUNTRY: USA
 ZIP: 22320
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/979,847
 FILING DATE: 26-NOV-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: BERRIDGE, WILLIAM P.
 REGISTRATION NUMBER: 30,024
 REFERENCE/DOCKET NUMBER: WPB 39046A
 TELEPHONE: 703-836-2787
 TELEFAX: 703-836-6400
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 297 base pairs
 TYPE: nucleotide
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-979-847-2
 Query Match 42.4%; Score 212.6; DB 7; Length 297;
 Best Local Similarity 94.0%; Pred. No. 1.4e-60;
 Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 Qy 1 CAATATCTCAGATTAATCAATGAGGCTGTTGCTTATAGCCAGTGTACCTAGCCCT 60
 Db 61 CAATACTCAGATTAATCAATGAGGCTGTTGCTTATAGCCAGTGTACCTAGCCCT 120
 Qy 61 TATATCTGCTTCCCAATATACAGAGGAGAGGTTTACAGTCTGAGCTTCAG 120
 Db 121 TATACAGTCTTCCCAATATACAGAGGAGAGGTTTACAGTCTGAGCTTCAG 180
 Qy 121 GATGCTTCTTCTGATCCCTGATACCTGACTCTCAATTTCTGTTGCTTGAAGAT 180
 Db 181 GATGCTTCTTCTGATCCCTGATACCTGACTCTCAATTTCTGTTGCTTGAAGAT 240

QY 181 ACTTCAACCAACATCTCACTGACATTTTACCCCAAGGTTTCAGG 235
DB 241 CCTTTGAACCAACGTCTCACTGACATTTTACCCCAAGGTTTCAGG 295

RESULT 14

US-09-864-761-15266/c
Sequence 15266, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 15266
LENGTH: 540
TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:
OTHER INFORMATION: MAP TO AC025746.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.69
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.73
US-09-864-761-15266

Query Match 42.3%; Score 212; DB 10; Length 540;
Best Local Similarity 71.1%; Pred. No. 2,9e-60;
Matches 324; Conservative 0; Mismatches 125; Indels 7; Gaps 3;
QY 53 CTAGCCCTTACTCTGCTTCCCAATACGAGAGAGAGAGAGGTTTACAGTCTGG 112

DB 540 CCAACCCCTTACCTGCTCTCTCAATATCCAGAGAGAGATGGTTACAGTTCTGG 481
QY 113 ACCTTGAAGATGCTTCTTCTGATCCCTGTATACCTCTGATCTTCTTGTGGCT 172
DB 480 AACTCAAGATGCTTCTTCTGATACCTGTACCTGTACCTCCAGCTTCTTGTGCT 421
QY 173 TTGAAGTACTTCAAAACCAACATCTCACTACCTGATATTTTACCCCA-GGGTTC 231
DB 420 TTGAAGTATCCCAACACATATCTCCACCTTACGTGATGATGCTTGGCCCAAGGGGTTT 361
QY 232 AGGATGATCCCATCTATTTGGCAAGGCTTAGGCCCAAGTTCAGCTCTCA-T 289
DB 360 AGGATGATCCCTTATCTGTTGGTCAAGGCTAGCCCAAGATCTAGGCTCTCACT 301
QY 290 ACCTGACATCTGCTTCCGTAGTGTATGATTTTCTTGGCCGCCCATTCAGAAACC 349
DB 300 CAGGACATCTGCTCTTCAATAGTATGATTTTCTTGGCTGCACTTGGAGAGCC 241
QY 350 TTGTGCATCAAGCCACCAAGGCTTTCAATTTCTCCCTACTGTGGCTACATGTT 409
DB 240 TCATGCACAGAGGCTACTTATGATCTCTGACCTTCTTACTATCAAGGGTCAAGTTG 181
QY 410 TCCAACCAAGGCTCACTGCTCAAGAGGT---TACTTAGGCTTAAATATTC 465
DB 180 TCTGGTGAAGTCCAGCTTTGCTTACAGAGGTCAATATCTTAGGCTTATTTAGCC 121
QY 466 AAAGGACCAAGGCTCTCACTGAGGAACATTCAG 501
DB 120 AGAGGACCAAGGCTCTCAAGGAATGATGATCAG 85

RESULT 15

US-08-979-847-8
Sequence 8, Application US/08979847
Publication No. US20030039664A1

GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNARD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787


```

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 645 base pairs
;   TYPE: nucleotide
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-979-847-8

```

```

Query Match      42.2%; Score 211.4; DB 7; Length 645;
Best Local Similarity 89.9%; Pred. No. Se-60;
Matches 250; Conservative 0; Mismatches 22; Indels 6; Gaps 2;

```

```

QY 230 TCAGGATAGTCCCATCTATTGGCCAGGATTAGCCCAAGACTTGAAGCCATCTCAT 289
Db 1 TCAGGATAGCCCCCATCTATTGGCCAGGATTAGCCCAAGACTTGAAGCCATCTCAT 60
QY 290 ACCTGGACA--CTTGTCTTGGGTAGTGATGATTACTTTGGCCGCCCATTCAGAAA 347
Db 61 ACCTGGACACTCTTGTCTTGGGTAGTGATGATTACTTTAGTCCGCCGTTTCAGAAA 120
QY 348 CCTGTGTCATCAAGGCCAAGCGCTCTCAATTTCTGCTACCTGTGGCTACATGG 407
Db 121 CCTGTGTCATCAAGGCCAAGCGCTCTTACTTCTCTCACTACCTGTGGCTACATGG 180
QY 408 TTTCCAACCAAAAGGCTCACTGTGCTCAGACAG---GTTACTTAGGGCTAAATTAT 463
Db 181 TTTCCAACCAAAAGGCTGTGGCTGTGCTCAGAGAGATTAGATACTNAGGGCTAAATTAT 240
QY 464 CCAAGGCAACAGGGGCTTCAAGTGAAGAAACATCCAG 501
Db 241 CCAAGGCAACAGGGGCTTCAAGTGAAGAAACATCCAG 278

```

Search completed: April 19, 2003, 14:55:23
 Job time : 80.717 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 11:45:23 ; Search time 1268.51 Seconds
(without alignments)
6396.429 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501
Sequence: 1 caagatctcagatcatca.....tcagtgcagacacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estbda: *
2: em_esthum: *
3: em_estnu: *
4: em_estnu: *
5: em_estnu: *
6: em_estnu: *
7: em_estnu: *
8: em_estnu: *
9: gb_est1: *
10: gb_est2: *
11: gb_est3: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estfun: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vit: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	384.2	76.7	521 17	AZ517418 RPCI-11-4
2	349	69.7	674 17	AQ111469 CIT-HSP-2
3	324.2	64.7	533 17	B17809 347L9.TVB
4	300	59.9	594 17	AQ193544 CIT-HSP-2
5	266.8	53.0	595 17	B59017 CIT-HSP-201
6	265.6	53.0	745 9	AU121943 AU121943

C 7	256.6	51.2	750 9	AL698185
C 8	253.2	50.5	291 12	BF989696
C 9	252.6	50.4	707 17	B66771
C 10	208	41.5	736 10	AV731083
C 11	206.2	41.2	537 17	AQ270891
C 12	204.4	40.8	416 17	AQ041889
C 13	204.4	40.8	518 17	AQ361037
C 14	203.8	40.7	434 10	BE168410
C 15	201	40.1	408 17	AQ033042
C 16	190	37.9	549 17	AQ231407
C 17	177.8	35.5	405 17	AQ033015
C 18	177.2	35.4	557 17	AQ061550
C 19	175.8	35.1	676 9	AL705035
C 20	168	33.5	518 17	AQ243044
C 21	164.8	32.9	428 17	AQ132081
C 22	161.6	32.3	442 17	AQ075734
C 23	159.2	31.8	541 17	AQ070212
C 24	157.2	31.4	491 17	B17135
C 25	155	30.9	431 17	AQ221075
C 26	152.4	30.4	442 17	AQ0701392
C 27	150.6	30.1	467 17	AQ0876467
C 28	148	29.5	456 17	AQ032635
C 29	148	29.5	519 17	AQ153050
C 30	147.8	29.5	449 17	AQ0701255
C 31	146	29.1	526 17	AQ217989
C 32	143	28.5	332 17	AQ024733
C 33	140.6	28.1	737 17	B67137
C 34	140.4	28.0	437 17	B47870
C 35	140.2	28.0	461 17	AQ225099
C 36	138.4	27.6	518 17	AQ0807540
C 37	132.6	26.5	518 17	AQ284525
C 38	132.2	26.4	429 17	AQ028930
C 39	131.4	26.2	431 17	AQ0736815
C 40	127.2	25.4	433 17	AQ145803
C 41	123	24.6	469 17	AQ223144
C 42	121	24.2	398 17	AQ044697
C 43	119.8	23.9	454 17	AQ226226
C 44	118.2	23.6	533 17	AQ479087
C 45			383 17	AQ192721

ALIGNMENTS

RESULT 1
AZ517418/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

521 bp DNA linear GSS 16-OCT-2000
RPCI-11-4708-TV RPCI-11 Homo sapiens genomic clone RPCI-11-4708,
DNA sequence.
AZ517418
AZ517418.1 GI:10826891
GSS.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
BAC end sequences of library RPCI-11
Unpublished (1997)
Other GSSs: RPCI11-4708-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igir.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet cs (info@reagen.com). BAC end search page:
http://www.tigr.org/cdb/hungen/bac_end_search.html.
This BAC end was generated during the Red process and may have
higher chance of clone tracking errors.

Seq primer: 17
Class: BAC ends.

FEATURES
source
location/Qualifiers

1..521
/organism="Homo sapiens"
/db_xref="GDB:7518007"
/db_xref="taxon:9606"
/clone="RPC1-11-4708"
/clone_1lb="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.5; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC library"
BASE COUNT 145 a 93 c 147 g 136 t
ORIGIN

Query Match 76.7%; Score 384.2; DB 17; Length 521;
Best Local Similarity 90.8%; Pred. No. 1e-111; Indels 6; Gaps 2;
Matches 433; Conservative 0; Mismatches 38; Indels 6; Gaps 2;

QY 31 GTTCTCTATACCAAGCTTACCTTACCTTATCTCTCTTCCCAATACAGAGAA 90
DB 521 GTTCTCTATACCAAGCTTACCTTATCTCTCTTCCCAATACAGAGAA 462
QY 91 GCAGAGTGTTCACAGTCTGACCTTCAAGATCCTTCTTGCACTCCCTGATCCT 150
DB 461 GCAGAGTGTTCACAGTCTGACCTTCAAGATCCTTCTTGCACTCCCTGATCCT 402
QY 151 GACTGCAATCTCTGTTGGCTTGAAGATCTCAACCAACATCTCACTGACCTG 210
DB 401 GACTGCAATCTCTGTTGGCTTGAAGATCTTGAACCAACATCTCACTGACCTG 342
QY 211 ACTATTTTACCCCAAGAGTTCAAGGATAGTCCCATCTATTGGCCAGCATTAAGCCAA 270
DB 341 ACTGTTTACCCCAAGAGTTCAAGGATAGTCCCATCTATTGGCCAGCATTAAGCCAA 282
QY 271 GACTTGAGCAATCTCTCACTGACCA--CTTGCTCTTGGTGGTGGATGATTTACTT 328
DB 281 GACTTGAGCAATCTCTCACTGACCACTTGTCTTCAATGAGGATGATTTAAAT 222
QY 329 TTGGCGCCCATTCGAACCTTTGSCATCAAGTCAACCAAGCGCTTCAATTTCCCTC 388
DB 221 TTGGCGCCCATTCGAACCTTTGSCATCAAGTCAACCAAGCGCTTCAATTTCCCTC 162
QY 389 GCTACCTGTGCTACATGTTTCAACCAAGGTTCAACTCTGCTCAGACAGAGT--- 444
DB 161 GCTACCTGTGCTACATGTTTCAACCAAGGTTCAACTCTGCTCAGACAGAGTTAA 102
QY 445 TACTTAGGGCTAAATTTATTCAAAGGACACAGGACCTCAGTGAGGAGACATCCAG 501
DB 101 TACTTAGGGCTAAATTTATTCAAAGGACACAGGACCTCAGTGAGGAGATCCAG 45

RESULT 2
AQ11469/c 674 bp DNA linear GSS 29-AUG-1998
LOCUS CIT-HSP-2372020.TR CIT-HSP Homo sapiens genomic clone 2372020, DNA
DEFINITION
ACCESSION AQ11469
VERSION AQ11469.1 GI:3488126
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 674)
Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher K., Golden K.,
Berry K., Granger D., Suh E., Wible C., Shizuya H., Simon M. and
Venter J.C.

TITLE
JOURNAL
COMMENT
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
Unpublished (1998)
Other GSSs: CIT-HSP-2372020.TP
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: madams@tigr.org
Clones are available from Research Genetics (info@reagen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
source
location/Qualifiers

1..674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2372020"
/clone_1lb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelBAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 180 a 128 c 184 g 182 t
ORIGIN

Query Match 69.7%; Score 349; DB 17; Length 674;
Best Local Similarity 90.3%; Pred. No. 2.1e-100; Indels 14; Gaps 3;
Matches 411; Conservative 0; Mismatches 30; Indels 14; Gaps 3;

QY 53 CTAGCCCTTATACCTGCTTCCCAATACCAAGAGAGAGTTCACATCGCGG 112
DB 674 CTAGCCCTTATACCTGCTTCCCAATACCAAGAGAGAGTTCACATCGCGG 615
QY 113 ACCCTCAGAGATCCTTCTTCTGCACTCCCTGATACATCTGACTCTCAATCTTGTGCT 172
DB 614 ACCTTAAGATGCTTCTTCTGCACTCCCTGATACATCTGACTCTCAATCTTGTGCT 555
QY 173 TTGAAGATCTTCAAAACCAATCTCACTGACCTGACATATTTTACCCAGGTTCA 232
DB 554 TTGAAGATCTTCTCAAAACCAATCTCACTGACCTGACATATTTTACCCAGGTTCA 495
QY 223 GGGATATGCCCATCTATTGGCCAGGATTAAGCCAGACTGAGCAATCTCATACC 292
DB 494 GGGATATGCCCATCTATTGGCCAGGATTAAGCCAGACTGAGCAATCTCATACC 435
QY 293 TGGACA--CTTGCTCTTGGTGGTGGATGATTTTACTTTTGGCCGCCATTGAGAACT 350
DB 434 AGGACATCTTGTCTTCTGACATGATGATTTTCTTTAGCGCGCTTTCAGAAACT 375
QY 351 TGTGCGATCAAGCCACCAAGGCTCTTCAATTTCTGCTACTGTGAGCTACATGTTT 410
DB 374 TGTGCGATCAAGCCACCAAGGCTCTTCAATTTCTGCTACTGTGAGCTACATGTTT 323
QY 411 CCAACCAAGGCTCAATCTGCTCAGACAGCT---TACTTAGGCTAAATATCA 466
DB 322 CCAACCAAGGCTCAATCTGCTCAGACAGGTTAAATATTAGGCTAAATATCA 263
QY 467 AAGGACCAAGGCTCTGATGAGAGACATCCAG 501
DB 262 AAGGACCAAGGCTCTGATGAGAGATGATCCAG 228

RESULT 3
B17809/c 533 bp DNA linear GSS 04-JUN-1998
LOCUS B17809
DEFINITION 347L9.TVB CIT978SK1 Homo sapiens genomic clone A-347L09, DNA
ACCESSION B17809
VERSION B17809.1 GI:2125558

KEYWORDS GSS.
SOURCE human.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 533)
AUTHORS Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.
TITLE Use of a BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@ligr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES
source location/Qualifiers
1..533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A-347L09"
/clone_1lb="CIT978SKA1"
/sex="Female"
/cell_type="Fibroblast"
/note="Vector: pBAC108L, Site_1: HindIII; Site_2: HindIII;
Caltech Human BAC Library A1"

BASE COUNT 154 a 97 c 144 g 138 t
ORIGIN

Query Match 64.7%; Score 324.2; DB 17; Length 533;
Best Local Similarity 86.0%; Pred. No. 1.6e-92;
Matches 398; Conservative 0; Mismatches 53; Indels 12; Gaps 3;

QY 50 TACGAGCCCTTACTCTGCTTCCCAATACGAGAGACAGTGGTTACAGTCC 109
DB 533 TACTTAACCTTAATCTGCTTCCCAATGCCAGAGAGGACAGATGATATAGTCT 474

QY 110 TGGACCTTCAGAGATGCTTCTTGTCATCCCTGTACATCTGACTCTCAATTCGTTTG 169
DB 473 TGGACCTTAAGAGATGCTTCTTGTCATCCCTGTACATCTGACTCTCAATTCGTTTG 414

QY 170 CCTTGAAGATCTTCAAAACCCACATCTCACTGACGATTTTACCCCAAGGCT 229
DB 413 CCTTGAAGATCTTCAAAACCCACATCTCACTGACGATTTTACCCCAAGGCT 354

QY 230 TCAGGATAGTCCCATCTATTGGCCAGGCACT-----TAGCCCAAGCTTGAGCCANT 283
DB 353 TCAGGATAGTCCCATCTATTGGCCAGGCACTCAACCCCAAGCTTGAGCCANT 294

QY 284 CCTCATCTGACACT--TGCTCTCGTAGTGTAGTATGATTTATTTGGCCGCAAT 341
DB 293 TCTATATCTGAACACTATGTTGTTGTATGATGATGATTTATCTAGCGCCCTG 234

QY 342 CAGAAACCTTGCCATCAGGCAACCCCAAGCGCTTCAATTTCTCGTACCTTGCT 401
DB 233 CAGAAACCTTGCCATCAGGCAACCCCAAGCGCTTCAATTTCTCGTACCTTGCT 174

QY 402 ACATGGTTCCAAACCAAGGCTCAACTGTGTCACAGAGGT-----TACTTAGGGCTAA 457
DB 173 ACATGGTTCCAAACCAAGGCTCAACTGTGTCATAGAGGTTAAATTAATCTTAGGGTTAA 114

QY 458 AATTATCCAAAGGCAACGAGGCTCTAGTGAACACATCCA 500
DB 113 AATTATCCAAAGGCAACGAGGCTCTAGGAGAAATGATCCA 71

RESULT 4
AQ193544/c

LOCUS AQ193544 594 bp DNA linear GSS 16-SEP-1998
DEFINITION CIT-HSP-2381P7.TR CIT-HSP Homo sapiens genomic clone 2381P7, DNA
sequence.
ACCESSION AQ193544
VERSION AQ193544.1 GI:3603659
KEYWORDS GSS.
SOURCE human.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 594)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Baas,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2381P7.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@ligr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reversee
Class: BAC ends.

FEATURES
source location/Qualifiers
1..594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2381P7"
/clone_1lb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelBAC11, Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 151 a 111 c 170 g 162 t
ORIGIN

Query Match 59.9%; Score 300; DB 17; Length 594;
Best Local Similarity 91.7%; Pred. No. 9.4e-85;
Matches 341; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

QY 136 ATCCCTGATACCTCTGACTCTCAATTTCTTTGCTTGAAGATCTTCAACCCACA 195
DB 594 ATCCCTGATACCTCTGACTCTCAATCTGTTTGCTTGAAGATCTTCAACCTAATG 535

QY 196 TCTCACTCAGCTGAGCACTTTTACCCCAAGGTTCAAGGATAGTCCCATCTATTGGC 255
DB 534 TCTCACTCAGCTGAGCACTTTTACCCCAAGGTTCAAGGATAGTCCCATCTATTGGC 475

QY 256 CAGGATTAAGCCCAAGACTTGAAGCAATCTCATACCTGAGCA--CTGTCTCTGGTAG 313
DB 474 CAGGATTAAGCCCAAGACTTGAAGCAATCTCATACCTGAGCACTGTCTCTGGTAG 415

QY 314 GTGATGATTTACTTTTGGCCGCCATTCAGAAAACCTTTGCCATGAAGCAACCCAGCG 373
DB 414 ATGATGATTTATTTTATTTTATTTAGTCCCGTTCAAAAACCTTTGCCATGAAGCAACCCAGCG 355

QY 374 CTCTTCATTTCTCTGCTGCTACCTGTGTGCTACATGTTTCCAAACCAAGGCTCAACTCTGC 433
DB 354 CTCTTCATTTCTCTGCTGCTACCTGTGTGCTACATGTTTCCAAACCAAGGCTCAACTCTGC 295

QY 434 TCACAGCAGGT-----TACTTAGGGCTAAATTTATCCAAAGGCAACGAGGCTCTAGTAG 489
DB 294 TCACAGCAGGTAAATTTAGGGCTAAATTTATCCAAAGGCAACGAGGCTCTAGTAG 235

QY 490 GAACATCCAG 501
|||||

DB	234	GAACGTTCCAG	223
RESULT 5			
LOCUS	B59017/c	595 bp	DNA linear GSS 20-JUN-1998
DEFINITION	CIT-HSP-2014K19.TR CIT-HSP Hom sapiens genomic clone 2014K19, DNA sequence.		
ACCESSION	B59017		
VERSION	B59017.1	GI:2613735	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 595) Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M., and Venter,J.C.		
TITLE	Use of a random BAC End Sequence Database for Sequence-Ready Map Building		
JOURNAL	Unpublished (1997)		
COMMENT	Other_GSSs: CIT-HSP-2014K19.TF Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mcadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html Seq primer: M13 Reverse Class: BAC ends.		
FEATURES	source	Location/Qualifiers	
		1..595	
		/organism="Homo sapiens"	
		/db_xref="GDB:7042821"	
		/db_xref="taxon:9606"	
		/clone="2014K19"	
		/clone_1fb="CIT-HSP"	
		/sex="Male"	
		/cell_type="Sperm"	
		/note="Vector: pBelovAC11; Site_1: HindIII; Site_2: HindIII"	
BASE COUNT	158 a 104 c 159 g 174 t		
ORIGIN			
Query Match	53.3%; Score 266.9; DB 17; Length 595;		
Best Local Similarity	89.2%; Pred. No. 3.9e-74;		
Matches 323; Conservative 0; Mismatches 32; Indels 7; Gaps 3			
QY	146 ATCTGAGCTCTCAATCTGTTGTTGCTTTGAGAGATCTTCAACCAACATCTCAACTCA	205	
DB	595 ATCTGAGCTCTCAATCTGTTGTTGCTTTGAGAGATCTTCAACCAACATCTCAACTCA	536	
QY	206 CCGGACATTTTAAACCCCAAGGTTGAGGATAGTCCCATCATTTGGCGAGGACTTAG	265	
DB	535 CCGGAGTTGTTTAAACCCCAAGGTTGAGGATAGTCCCATCATTTGGCGAGGACTTAG	476	
QY	266 CCCAAGACTTGAAGCAATCTCTATACCTGAGACAT--TGCTTCGGTAGTGATGATT	323	
DB	475 CTCAGACTTGAAGCAATCTCTATACCTGAGACATCTCTCTTGACGATGATGATT	416	
QY	324 TACTTTGGCCGCCATTTCAAGAACTTTGTGCCATCAAGCACCCAGGCTCTTAATT	383	
DB	415 TACTTTTACTGCTCTTTTCAAGAACTTTGTGCCATCAAGCACCCAGGCTCTTAATT	356	
QY	384 TCTCGCTACTGTGCTCATGTTTCCAAACCAAGGCTCAACTCTGTACACAGCAG	443	
DB	355 TCTCTCCACCTGT-CATTTAAGTTTCCAAACCAAGGATCACTCTGTGTACACAGTAG	297	

QY	444	T-----TACTTATGGGCTAAATATATATCAAAAGGACACGAGGCCCTCAGTGAAGAACATCC	499
Db	236	TTAAATACTTAAGTGTAAATATATATCAAAAGGACACGAGGCCCTCAGTGAAGAACATATCC	237
QY	500	AG 501	
Db	236	AG 235	
RESULT 6			
LOCUS	AU121943	745 bp	mRNA linear EST 01-AUG-20020
DEFINITION	AU121943 MAMMAL Homo sapiens cdna clone MAMMAL001324 5', mRNA		
ACCESSION	AU121943		
VERSION	AU121943		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 745) Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isozaki, T.		
TITLE	HRI human CDNA project		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Takao Iwogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human CDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.		
FEATURES			
Source	Location/Qualifiers 1..745 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="MAMMAL001324" /clone_lib="MAMMAL" /tissue_type="mammary gland" /note="Vector: pMEB5FL3"		
BASE COUNT	214 a 196 c 156 g 176 t	3 others	
ORIGIN			
Query Match	53.0%, Score 265.6, DB 9, Length 745,		
Best Local Similarity	89.8%, Pred. No. 1.1e-73;		
Matches 309; Conservative 0; Mismatches 29; Indels 6; Gaps 2;			
QY	164	TTGTTGCTTTGAATATCTCAAAACCCAAACATCTCACTCACTGAGACTATTTTACCCC	223
Db	1	TTGTTGCTTTGAATATCTTTCAAAACCCAAATCTCAATTCACCTGAGACTGTTTACCCC	60
QY	224	ANGGTTAGGAGATGTGCCCATTAATTGGCAGAGCATTTAGCCCAAGACTTGAGCCAAAT	283
Db	61	ANGGTTAGGAGAGAGCCCATCTAATTTGGCAGAGCATTTAGCCCAAGACTTGAGCCAAAT	120
QY	284	CCTCATACCTGAGCA - CTGTCTCTTGGGTAGTGATGATTTACTTTTGGCCGCCATT	341
Db	121	TCTCACTACCTGAGCACTGTGTCTTCACTGATCATGATGATTTACTTTTGGCCGCCATT	180
QY	342	CAGAAACCTTGTGCATCAAGCACCAGAGGCTTTCAATTTCTCTGCTACCTGTGCT	401
Db	181	CAGAAACCTTGTGCATCAAGCACCAGAGGCTTTCAATTTCTCTGCTACCTGTGCT	240
QY	402	ACATGATTTCCAAACCAAGGCTCAACTGTGCTACAGACAG - - - TTACTTAGGCTAA	457
Db	241	ACCATGATTTCCAAACCAAGGCTCAACTGTGCTACAGAGGCTAAATCTTAGGCTTA	300

Qy 458 AATTATCCAAAGGACGAGGCGCTCAGTGAAGAAACATCCAG 501
Db 301 AATTATCTTAAGGACGAGGCGCTCAGTGAAGATGATACCCAG 344

RESULT 7
AL698185/c 750 bp mRNA linear EST 21-MAR-2002
LOCUS DKFZP686N20106.F1 686 (synonym: hlc3) Homo sapiens cDNA clone
DEFINITION DKFZP686N20106.5, mRNA sequence.
ACCESSION AL698185
VERSION AL698185.1 GI:19618725
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 750)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (2001)
COMMENT Contact: Ottenwaelder B
MIPS
Am Klopferstr. 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de
Sequenced by MediGenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. No sl sequence
available.
This clone (DKFZP686N20106) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcentrum Heubnerweg 6, 14059
Berlin-Charlottenburg; GERMANY; Email: clone@rzpd.de.

FEATURES
source
1..750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP686N20106"
/clone_lib="686 (synonym: hlc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SfilA; Site_2: SfilB;
cDNA-collection"

BASE COUNT 195 a 151 c 194 g 210 t

ORIGIN
Query Match 51.2%; Score 256.6; DB 9; Length 750;
Best Local Similarity 73.1%; Pred. No. 8.2e-71;
Matches 372; Conservative 0; Mismatches 129; Indels 8; Gaps 3;

Qy 1 CAAGATCTCAGGATTAATCATGAGGCTGTTCTCTATATAGCAGCTGTACT--AGCC 58
Db 582 CAAGATCTTAAGCTATATCATGAGGAGTAATTCCTATATAGCAGCTGTATCCCAACC 523
Qy 59 CTTAATCTCTGTTTCCCAATATCAAGAGAGAGAGAGGTTTACAGTCTCGACCTTC 118
Db 522 CCTAATACCTGCTTCTCTCAATATACAGAGAGAGAGAGGTTTCTGAGACCTCA 463
Qy 119 AGGATGCGCTTCTGCAATCCGTATACCTGACCTCAATTTCTGTTGCTTTGAAG 178
Db 462 AGGATGCGCTTCTGCAATCCGTATATGAGGCTTCTGCTTCTGCTTTGAGG 403
Qy 179 AATCTTCAAAACCAATCTCAACTCACTGTGACTATTTTACCCCAAGGTTCAAGGATA 238
Db 402 ATCCACAGACACACAGCTGCCAATTAATGATGACGGTCTTCAAGGTTAGGATA 343
Qy 239 GTCCCATCTATTTGGCCGAGCATTAAGCCAGACTGAGCCAACTCTTA--TACCTGA 296
Db 342 GCCATATCTGTGTGTGTCAGGACATGAGCCCAAGATCTAGGCCACTTCTCAAGTCCAGCA 283

Qy 297 CACTTGTCTTCGAGTAGTGATGATTTACTTTTGGCCGCCATTACAGAAACCTTGTGCC 356
Db 282 CTGTGTCTCTCAATATGATGATATTTACTTTGGCTACAGGATTCAGAACTTCATGCC 223
Qy 357 ATCAAGCCACCCAGGCGCTTTCATTTTCCCTGCTACTGTGGCTACATGTTTCCAAAC 416
Db 222 AGCAGGCTACTCTGATATCACTTGAACCTTTTGAATCAAGGGATGACGGCTCTAAT 163
Qy 417 CAAGGCTCACTCTGCTCACAGCAGT---TACTTAGGCTAAATATTCACAAAGCA 472
Db 162 TGAGGCCCACTCTGCTTACACAGTCAATATCTTAGGCTTAATCTTACCCAGAGAA 103
Qy 473 CCAGGCGCTCAGTGAAGAACATCCAG 501
Db 102 CCAACACCTCAGCAAGATGAAGACAG 74

RESULT 8
BF989696/c 291 bp mRNA linear EST 23-JAN-2001
LOCUS MR2-GN0127-231000-004-e11 GN0127 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF989696
ACCESSION BF989696
VERSION BF989696.1 GI:12396021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 291)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&t=MR2-GN0127-
231000-004-e11&t3=2000-10-23&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 290.
Location/Qualifiers
1..291
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0127"
/dev_stage="Adult"
/note="Organ: placenta, normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORBESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT 86 a 54 c 82 g 68 t

ORIGIN
Query Match 50.5%; Score 253.2; DB 12; Length 291;
Best Local Similarity 96.4%; Pred. No. 6.1e-70;
Matches 269; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY	4	GACACAGATTATCAAAAGAGGCTGTGTTGCTCTAAATGACGAGCTGPAACCTGAT	63
Db	280	GAACCTTAGGATTATCAATAGAGGCTGTGTTGCTCTAAATGACGAGCTGPAACCTGAT	221
QY	64	ACTGCTGTTTTCCCAATAACGAGAGGACAGTGGTTTACAGTCTGACCTTGAGAT	123
Db	220	ACTGCTGTTTTCCCAATATACGAGAGGACAGTGGTTTACAGTCTGACCTTGAGAT	161
QY	124	GCGTCTCTGTGATCCCTGTATCACTCTCAATCTGTGTGGCGTTGAAATGACT	183
Db	160	GCGTCTCTGTGATCCCTGTATCACTCTGTGACTCTCAATCTGTGTGTGGCGTTGAAATGACT	101
QY	184	TCAAACCCAACTCAACTCACTGAGATATTTAACCCCAAGGTTGACGAGTATGTCCTC	243
Db	100	TCAAACCCAACTCAACTCACTGAGATATTTAACCCCAAGGTTGACGAGTATGTCCTC	41
QY	244	CATCTATTTGGCAGGACATTAGCCCAAGACTGAGCCAA	282
Db	40	CATCTATTTT-GCAGGACATTAGCCTTAAGTCAAGNCCA	3

RESULT 9

LOCUS	B66771	707 bp	DNA	linear	GSS 21-JUN-1998
DEFINITION	Citr-HSP-2015D21.TFB Cit-HSP Homc; sapiens genomic clone 2015D21, DNA				
ACCESSION	sequence.				
VERSION	B66771				
KEYWORDS	B66771.1	GI:2640749			
SOURCE	GSS.				
	human.				

ORGANISM

REFERENCE
AUTHORS
Adams, M.D., Rounsley, S.D., Field, C.E., Baas, S., Linher, K., Golden

AUTHORS

TITLE	K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
JOURNAL	Use of a random BAC End Sequence Database for Sequence-Ready Map Building
COMMENT	Unpublished (1997) Other GSSs: CIT-HSP-2015D21.TR

COMMENT

Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page: http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.

FEATURES
Source

```

FEATURES
source
Location/Qualifiers
1..707
/organism="Homo sapiens"
/db_xref="GDB:704339"
/db_xref="RefSeq:9606"
/clone="2015D21"
/clone_id="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOAc11; Site_1: HindIII; Site_2: HindIII"

```

BASE COUNT
ORIGIN

BASE COUNT	195 a	183 c	144 g	165 t
ORIGIN				

Query Match

Query Match	50.4%	Score 252.6	DB 17	Length 707
Best Local Similarity	71.4%	Pred. No. 1.5e-69		
Matches 362	Conservative	0	Mismatches 139	Indels 6
				Gaps 2

Qy

1 CAAGATCTCAGGATTATCAATGAGCGTGTGTTCCCTATAGCCAGCTGTACCTAGCCCT 60

Db	176	CAAGATCTTAGACTCATTCAATGAGGACGATTAATCTCTAATATCCAGCTGATCCAAACCCC	235
Qy	61	TATATCTGCTTTTCCCAATAACGAGAGAGACGAGGTGTTTCAAGCTCCGACCTTCAG	120
Db	236	TATATCCCTGCCCCCTTCAAAATATACGAGAGGACGAAATGGTCTACTGTTCTGGATTCGAG	295
Qy	121	GATACCTCTTCTTGACATCCCTGTACATCCAGACTCTCAATTCCTTGTGTTCCTTGAAGAT	180
Db	296	GATACCTTCTCTTGCAATGTGTTGCACTGTGATCCCAAGTTTCTTGTGCTTTGAAGAT	355
Qy	181	ACTTCAAACCCACATCTCAACTCACTGAGCTATTTTACCCCAAGGCTTCAGGATAGT	240
Db	356	CCCGAGAACCCCATCCAACTTACATGACCGCTCTTGCTCAAGGGTTTAGAGATAGC	415
Qy	241	CCCCATCATTTTGGCCAGGCATTTAGCCCAAGACTTGAGCAATCTCTCA--TACCTGGACA	298
Db	416	CTCATCCGTTTGGTACAGCACTGCGCTTAAGATCTTAGCCACTTCTCAATCCAGGATT	475
Qy	299	CTTGCTCTTGCGTAGTGATGATTTACTTTTGGCCGCCCATTCAGAACTTGTGCCAT	358
Db	476	CTGATGCTTCGGTATGTGAGATGATTTACTTTTGTCTACAGTTCCAGAGCCCTATACCG	535
Qy	359	CAAGCAACCCAAAGGCTCTCAATTCCTCGGTACACGTGGGTACATGTTTCCAAACCA	418
Db	536	CAGGTTACTTGATCTCTTGAACTTTCTAGCTATATAAGGGTATCAAGAGGTCTCAATCA	595
Qy	419	AAGGCTCAACTCTGCTACAGCAGGT---TACTTAGGGCTAAATATTATCCAAAGCACC	474
Db	596	AAGGCCAGGCTCTGCTCAACAACAAGTCAATATCTAGGCATTAATCTTAGCCAAAGAGGCC	655
Qy	475	AAGGCCCTCAAGTAGGAACATCCAG	501
Db	656	ATGGCTTCATCAAGGAAGGATTCAG	662

RESULT 10

RESULT 10

LOCUS	AV731083/c	736 bp	mRNA	linear	EST 17-OCT-2000
DEFINITION	AV731083	HTF Homo sapiens	CDNA clone HTFAH01	5'	mRNA sequence.
ACCESSION	AV731083				
VERSION	AV731083.1	GI:10840504			
KEYWORDS	EST.				
SOURCE	human.				

SOURCE

ORGANISM	Homosapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	
1 (bases 1 to 736)	

REFERENCE

TITLE Homo sapiens CDNA HTF clones
 JOURNAL Unpublished (2000)
 COMMENT Contact: Zeguang Han

COMMENT

Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel.: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
SOURCE

FEATURES	source	Location/Qualifiers
	1. .736	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="HTFAH01"
		/clone_id="HTF"
		/tissue_type="Hypothalamus"
		/dev_stage="Adult"
		/lab_host="SOLR"
		/note="Vector: pBluescript sk(-), Site_1: EcoRI, Site_2: XhoI"
BASE COUNT	193 a 147 c 185 g 211 t	

ORIGIN

Query Match 41.5%; Score 208; DB 10; Length 736;
Best Local Similarity 74.9%; Pred. No. 2.9e-55;
Matches 274; Conservative 0; Mismatches 90; Indels 2; Gaps 1;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTTGTTCTTATAGCCAGCTGTACCTAGCCCT 60
DB 376 CAAGATCTTACATCATCATGAGGAGGATTCCTTATATCCAGTTGTACCAACCCC 317
QY 61 TATATCTGCTTCCCAATATACAGAGAGAGAGTGTATACAGTCTGTACCTTACG 120
DB 316 TATATCCCTGCTCTCAATATCAAGAGAGAGAGTGTATACAGTCTGTACCTTACG 257
QY 121 GATGCTTCTTGTGATCCCTGTATCATCTGATCTCAATCTTGTGCTTGAAGAT 180
DB 256 GATGCTGCTTGTGCTTCTTCTTCTGCTGATCTGATCTGATCTTGTGCTTGAAGAT 197
QY 181 ACTTCAACCCCAACATCTCACTCACTGATCTTATTTTACCCCAAGGTTCAAGGATAGT 240
DB 196 CCAACAGACACATCACTCACTTGTGATGCTTGTGCTTGTGCTTGTGATGATGCT 137
QY 241 CCCATCTATTGTCAGGAGGATTTAGCCCAAGCTTGAAGCTTCA--TACTGTGACA 298
DB 136 CCTCATCTGTTGTGTGAGGCTGCTGCTCAAGATCTAGGCTCACTTCTTAAAGTCCAGGCACT 77
QY 299 CTGTGCTTGTGATGATGATGATTTACTTTTGTGCTGCTTCAAGAACTTGTGCTAT 358
DB 76 CTGTGCTTGTGATGATGATGATTTACTTATGCTTCACTTCAAGAACTTGTGCTAT 17
QY 359 CAAGCC 364
DB 16 CGTGCC 11

RESULT 11
LOCUS A0270891/c 537 bp DNA linear GSS 03-NOV-1998
DEFINITION HS 2047_A1_H02_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2047 Col=3 Row=O, DNA sequence.
ACCESSION A0270891
VERSION A0270891.1 GI:3823431
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 537)
Mahaitsas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L. Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
CONTACT: Mahaitsas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2047 Row: O Column: 3
Class: BAC ends
High quality sequence stop: 537.
Location/Qualifiers
1..537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2047 Col=3 Row=O"
/clone_1fb="CIT Approved Human Genomic Sperm Library D"
/sex="male"

FEATURES
SOURCE

BASE COUNT 125 a 97 c 146 g 161 t 8 others
ORIGIN

Query Match 41.2%; Score 206.2; DB 17; Length 537;
Best Local Similarity 90.0%; Pred. No. 9.2e-55;
Matches 243; Conservative 0; Mismatches 21; Indels 6; Gaps 2;

QY 238 AGTCCCATCTATTGTCAGGAGGATTTAGCCCAAGCTTGAAGCAATCTCATCTGAGAC 297
DB 536 AGCCCCATCTATTGTCAGGAGGATTTAGCCCAAGCTTGAAGCAATCTCTACTGAGAC 477
QY 298 A--CTTGTCTTGTGATGATGATTTACTTTTGTGCTGCTTCAAGAACTTGTGCT 355
DB 476 ACTGTGCTTGTGATGATGATTTACTTTTGTGCTGCTTCAAGAACTTGTGCTGCT 417
QY 356 CATCAAGCCCAAGGCTTCTTCAATTTCTGCTTCACTGCTGCTTCAAGGATTTCCAA 415
DB 416 CATCAAGCCCAAGGCTTCTTCAATTTCTGCTTCACTGCTGCTTCAAGGATTTCCAA 357
QY 416 CCAAGGCTCACTGCTTCTGCTTCAAGGATTTAAATTTAGGCTTAAATTTCAAGGCT 471
DB 356 CCAAGGCTCACTGCTTCTGCTTCAAGGATTTAAATTTAGGCTTAAATTTCAAGGCT 297
QY 472 ACCAGGCTTCACTGCTTCAAGGATTTAAATTTAGGCTTAAATTTCAAGGCT 501
DB 296 ACCAGGCTTCACTGCTTCAAGGATTTAAATTTAGGCTTAAATTTCAAGGCT 267

RESULT 12
LOCUS A0041889/c 416 bp DNA linear GSS 14-JUL-1998
DEFINITION CIT-HSP-2335P3.TR CIT-HSP Homo sapiens genomic clone 2335P3, DNA
sequence.
ACCESSION A0041889
VERSION A0041889.1 GI:3311083
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 416)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.B., Bass,S., Linher,K.,
Simon,M. and Venter,J.C. Granger,D., Suh,E., Wible,C., Shizuya,H.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C. Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
CONTACT: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@igf.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.figr.org/tcd/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers
1..416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2335P3"
/clone_1fb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11, Site_1: HindIII, Site_2:
HindIII"

FEATURES
SOURCE

BASE COUNT 119 a 85 c 113 g 99 t

ORIGIN

Seq primer: puc 18 forward
High quality sequence stop: 432.
Location/Qualifiers

FEATURES

source

1..434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="HT0513"
/dev_stage="Adult"
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 125 a 92 c 120 g 97 t
ORIGIN

Query Match 40.7%; Score 203.8; DB 10; Length 434;
Best Local Similarity 73.3%; Pred. No. 4.8e-54;
Matches 288; Conservative 0; Mismatches 102; Indels 3; Gaps 2;

QY 1 CAAGATCTCAGAGTATATATAGAGGCTTTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
DB |||||
413 CAAGATCTTACATCATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 354
QY TATACCTGCTTCCCAATATACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
DB |||||
353 G-TACTGCTCTCTCAATATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 295
QY 121 GATGCTTCTTCTGCATCCTGTATCATCTGATCTCAATCTTGTTCCTTGAAGAT 180
DB |||||
294 GATGCTTCTTCTGCATCCTGTATCATCTGATCTCAATCTTGTTCCTTGAAGAT 235
QY 181 ACTTCAACCAATCATCTCACTGAGTATTTTACCCCAAGGTTGAGGATGT 240
DB |||||
234 CCCACAGCCACATGTCCCACTTACGTCAGTCTTCCCAAGGCTTGAAGATGTC 175
QY 241 CCCCATCTTATTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 298
DB |||||
174 CTTACCTCTTGTGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 115
QY 299 CTTGCTCTTGGTATGATGATTTTCTTTGGCCGCGCATTCAGAACTTGTGCGAT 358
DB |||||
114 CTGCTCTTCAATATGATGATTTTCTTTGGCTACAGTTGGGAGGAGGAGGAGGAGGAGG 55
QY 359 CAAGCCACCCAGGCGCTTTCATTTCTCTGCT 391
DB |||||
54 CAGGCTACTCTAGATCTCTGAACCTTCTAGCT 22

RESULT 15 A0033042 408 bp DNA linear GSS 02-JUL-1998
LOCUS HS 2226 A1 D12 MF CIT Approved Human Genomic Sperm Library D Homo

DEFINITION sapiens genomic clone Plate=2226 Col=23 Row=G, DNA sequence.
ACCESSION A0033042
VERSION A0033042.1 GI:3285230

KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euleleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 408)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J.J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT High Throughput Sequencing Center

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2226 row: G column: 23
Class: BAC ends
High quality sequence stop: 408.
Location/Qualifiers

FEATURES

source

1..408
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=2226 Col=23 Row=G"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 115 a 83 c 105 g 98 t 7 others
ORIGIN

Query Match 40.1%; Score 201; DB 17; Length 408;
Best Local Similarity 69.5%; Pred. No. 3.7e-53;
Matches 282; Conservative 0; Mismatches 122; Indels 2; Gaps 1;

QY 43 CCAGCTGATACCTAGCCCTTATCTGCTTCCCAATATACAGAGGAGGAGGAGGAGGAGG 102
DB |||||
408 CCAATTTACCAATCATATACATCTCTCTCAATATCAAGAGGAGGAGGAGGAGGAGG 349
QY 103 ACAGTCTGACCTTCAAGATGCTTCTTGTGATCCCTGTACATCTGATCTCAATTC 162
DB |||||
348 ACTGTTTGGACNTGAAGATGCTTCTTGTGATCCCAATGACATCTGCTGCGCAGATT 289
QY 163 TTGTTGCTTGAAGATCTTCAACCAATCTCACTGATGATTTTACC 222
DB |||||
288 CTCTTGTCTTGAAGATCTTCAACCAATCTCACTGATGATTTTACC 229
QY 223 CAAGGTTTCAAGGATATCCCATCTATTTGGCCAGGAGGAGGAGGAGGAGGAGGAGG 282
DB |||||
228 CAGGAGTTTGAAGATATCCCATCTGTTTGTGATATTTGGCCAGGAGGAGGAGGAGG 169
QY 283 TCCTCA--TACCTGAGACATTTGCTTGGTATGATGATTTTGGCCGCGCAT 340
DB |||||
168 TTCTCAAGTCCAGGACATCTGATCTTCAATGATGATGATTTTGGCTACAGTT 109
QY 341 TCAGAAACCTTGTGCTATCAAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 400
DB |||||
108 TGAAGCTTTCATGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 49
QY 401 TACATGTTTCCAAACCAAGGCTCACTGCTACAGAGGTTA 446
DB |||||
48 TACAAAGATTTAAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3

Search completed: April 19, 2003, 14:46:34
Job time: 1274.51 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:28:30 ; Search time 1703.47 Seconds
(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500
Perfect score: 501
Sequence: 1 caagatcctcagatcatca.....tcagtgaggaacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Genembl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_com.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_scs.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_ay.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	10499	6 AX007980	AX007980 Sequence
2	501	100.0	56093	6 AX329572	AX329572 Sequence
3	501	100.0	56093	9 HSAC000064	AC000064 Human BAC
4	501	100.0	149194	9 AC007566	AC007566 Homo sapi
5	474	94.6	2938	6 AX000957	AX000957 Sequence
6	474	94.6	2938	6 AX027471	AX027471 Sequence
7	474	94.6	2938	6 AF072499	AF072499 Homo sapi
8	442	88.2	7582	6 AX000966	AX000966 Sequence
9	442	88.2	7582	6 AX027480	AX027480 Sequence
10	435	86.8	40205	6 AF045450	AF045450 Homo sapi
11	435	86.8	142742	9 AF121782	AF121782 Homo sapi
12	435	86.8	340000	9 HS21C080	AL163280 Homo sapi
13	428.6	85.5	8339	9 AL162912	AL162912 Human DNA
14	428.4	85.5	3372	6 AX000964	AX000964 Sequence
15	428.4	85.5	3372	6 AX027478	AX027478 Sequence
16	428.4	85.5	3372	6 AF072502	AF072502 Homo sapi
17	425.4	84.9	146545	2 AC104009	AC104009 Homo sapi
18	423.8	84.6	143063	9 AC040936	AC040936 Homo sapi
19	423.8	84.6	163166	9 AL356632	AL356632 Human DNA
20	423.8	84.6	176773	2 AC009867	AC009867 Homo sapi
21	422.2	84.3	2304	14 AF009668	AF009668 Multiple
22	422.2	84.3	108232	9 AP000654	AP000654 Homo sapi
23	422.2	84.3	156938	9 AL159163	AL159163 Human DNA
24	422.2	84.3	184523	2 AP002890	AP002890 Homo sapi
25	421.2	84.1	139744	9 AL133513	AL133513 Human DNA
26	420.6	84.0	132981	9 HS82311	Z83850 Human DNA s
27	420.6	84.0	161771	9 CNS01DRX	AL121579 Human chr
28	420.6	84.0	178152	2 AC108746	AC108746 Homo sapi
29	420.6	84.0	191426	2 AC026086	AC026086 Homo sapi
30	420.6	84.0	194545	2 AC087482	AC087482 Homo sapi
31	420.6	84.0	194718	2 AC026977	AC026977 Homo sapi
32	420.6	84.0	220807	9 AC093334	AC093334 Homo sapi
33	420.6	84.0	251124	9 HOAE000660	AE000660 Homo sapi
34	419.6	83.8	1158	6 A46012	A46012 Sequence 1
35	419.6	83.8	1158	6 A57048	A57048 Sequence 1
36	419.6	83.8	1158	6 A60113	A60113 Sequence 1
37	419.6	83.8	1158	6 A79463	A79463 Sequence 1
38	419.6	83.8	1158	6 A80197	A80197 Sequence 1
39	419.6	83.8	1158	6 A80306	A80306 Sequence 1
40	419.6	83.8	1158	6 AR035349	AR035349 Sequence
41	419.6	83.8	1158	6 AR036190	AR036190 Sequence
42	419.6	83.8	1158	6 AR036710	AR036710 Sequence
43	419.6	83.8	1158	6 AR076972	AR076972 Sequence
44	419.6	83.8	1158	6 AR094693	AR094693 Sequence
45	419.6	83.8	1158	6 AR183928	AR183928 Sequence

ALIGNMENTS

RESULT 1	AX007980	10499 bp	DNA	linear	PAT 06-SEP-2000
LOCUS	AX007980	Sequence 3 from Patent WO9967395.			
DEFINITION	AX007980				
ACCESSION	AX007980.1	GI:9995677			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 10499)
Perin,J.P., Rieger,F. and Alliel,P.M.
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses

JOURNAL

Patent: WO 9667395-A 3 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)

FEATURES

source

1.10499
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 2495 t
ORIGIN

Query Match 100.0%; Score 501.; DB 6; Length 10499;

Best Local Similarity 100.0%; Pred. No 2,5e-138; Mismatches 0; Indels 0; Gaps 0;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 CAAGATCTCAGGATTATCATGAGGCTGTGTTCTCTATACCCAGCTGTACCTAGCCCT 60
DB 5000 CAAGATCTCAGGATTATCATGAGGCTGTGTTCTCTATACCCAGCTGTACCTAGCCCT 5059
OY 61 TATACCTGCTTTCCCAATACCAAGAGAGCAAGTGGTTTACAGTCTTGAGCCTTCAG 120
DB 5060 TATACCTGCTTTCCCAATACCAAGAGAGCAAGTGGTTTACAGTCTTGAGCCTTCAG 5119
OY 121 GATGCTTCTTGATGATCCCTGTACATCTGTGACTCTGATTTCTGTTTCCCTTTGAAGT 180
DB 5120 GATGCTTCTTGATGATCCCTGTACATCTGTGACTCTGATTTCTGTTTCCCTTTGAAGT 5179
OY 181 ACTTCAACCCAACTCTCACTGACCTGACCTATTTTACCCCAAGGGTTTCAAGGATAGT 240
DB 5180 ACTTCAACCCAACTCTCACTGACCTGACCTATTTTACCCCAAGGGTTTCAAGGATAGT 5239
OY 241 CCCCATCTATTGTCGAGGAGGATTAAGCCCAAGACTTGAGCCATCTCATACCTGAGACT 300
DB 5240 CCCCATCTATTGTCGAGGAGGATTAAGCCCAAGACTTGAGCCATCTCATACCTGAGACT 5299
OY 301 TGTCTTGGTGTAGTGATGATTTACTTTTGCCGCCCTTGAGAAACCTTTGGCATCA 360
DB 5300 TGTCTTGGTGTAGTGATGATTTACTTTTGCCGCCCTTGAGAAACCTTTGGCATCA 5359
OY 361 AGCCACCAGAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCAA 420
DB 5360 AGCCACCAGAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCAA 5419
OY 421 GGCTCAACTCTGCTCAACAGAGGTTACTTGGGCTTAAATTTACAAAGGACACAGGGCC 480
DB 5420 GGCTCAACTCTGCTCAACAGAGGTTACTTGGGCTTAAATTTACAAAGGACACAGGGCC 5479
OY 481 CTCAGTGAGGAACACATCCAG 501
DB 5480 CTCAGTGAGGAACACATCCAG 5500
```

RESULT 2

AX329572 56093 bp DNA linear PAT 09-JAN-2002

LOCUS AX329572 Sequence 81 from Patent WO0194629.

ACCESSION AX329572

VERSION AX329572.1 GI:18102550

KEYWORDS

SOURCE

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

1 Young, P.E., Augustus, M., Carter, J.C., Ehner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.,
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 81 13-DEC-2001;

TITLE

Avalon Pharmaceuticals (US)

JOURNAL

FEATURES

source

1.56093
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Score 501.; DB 6; Length 56093;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 CAAGATCTCAGGATTATCATGAGGCTGTGTTCTCTATACCCAGCTGTACCTAGCCCT 60
DB 33000 CAAGATCTCAGGATTATCATGAGGCTGTGTTCTCTATACCCAGCTGTACCTAGCCCT 33059
OY 61 TATACCTGCTTTCCCAATACCAAGAGAGCAAGTGGTTTACAGTCTTGAGCCTTCAG 120
DB 33060 TATACCTGCTTTCCCAATACCAAGAGAGCAAGTGGTTTACAGTCTTGAGCCTTCAG 33119
OY 121 GATGCTTCTTGATGATCCCTGTACATCTGTGACTCTGATTTCTGTTTCCCTTTGAAGT 180
DB 33120 GATGCTTCTTGATGATCCCTGTACATCTGTGACTCTGATTTCTGTTTCCCTTTGAAGT 33179
OY 181 ACTTCAACCCAACTCTCACTGACCTGACCTATTTTACCCCAAGGGTTTCAAGGATAGT 240
DB 33180 ACTTCAACCCAACTCTCACTGACCTGACCTATTTTACCCCAAGGGTTTCAAGGATAGT 33239
OY 241 CCCCATCTATTGTCGAGGAGGATTAAGCCCAAGACTTGAGCCATCTCATACCTGAGACT 300
DB 33240 CCCCATCTATTGTCGAGGAGGATTAAGCCCAAGACTTGAGCCATCTCATACCTGAGACT 33299
OY 301 TGTCTTGGTGTAGTGATGATTTACTTTTGCCGCCCTTGAGAAACCTTTGGCATCA 360
DB 33300 TGTCTTGGTGTAGTGATGATTTACTTTTGCCGCCCTTGAGAAACCTTTGGCATCA 33359
OY 361 AGCCACCAGAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCAA 420
DB 33360 AGCCACCAGAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCAA 33419
OY 421 GGCTCAACTCTGCTCAACAGAGGTTACTTGGGCTTAAATTTACAAAGGACACAGGGCC 480
DB 33420 GGCTCAACTCTGCTCAACAGAGGTTACTTGGGCTTAAATTTACAAAGGACACAGGGCC 33479
OY 481 CTCAGTGAGGAACACATCCAG 501
DB 33480 CTCAGTGAGGAACACATCCAG 33500
```

RESULT 3

HSAC000064 56093 bp DNA linear PRI 13-NOV-1996

LOCUS

HSAC000064 Human BAC clone RG083M05 from 7q21-7q22, complete sequence.

ACCESSION

AC000064.1 GI:1669369

VERSION

KEYWORDS

HTG.

SOURCE

ORGANISM

Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

The sequence of H. sapiens BAC clone RG083M05
1 (bases 1 to 56093)
2 (bases 1 to 56093)
Unpublished (1996)
Waterston, R.
Direct Submission
Submitted (13-NOV-1996)
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
e-mail: sapiens@wustl.edu

NOTICE: This sequence may not represent the entire insert of this
sections once, or longer because we only sequence overlapping
neighborings submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The
library contains cloned DNA from a human male fibroblast cell line
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBelo

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this
clone is at base position 1 of H_RG083M05; actual end is at 56093
of H_RG083M05

This clone contains STS SW851725.

FEATURES

SOURCE

Location/Qualifiers

1. 56093

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone="H_RG083M05"

/clone_1fb="CTTB-978SK-B"

/complement(838..1131)

/rpt_family="ALU"

<1360..16971

/gene="WUGSC:H_RG083M05.1"

join(<1360..1503,4181..4370,4587..4774,6422..6556,

9483..9547,11631..11773,11864..12021,13131..13296,

14885..14988,16349..16546,16837..16971)

/gene="WUGSC:H_RG083M05.1"

/note="ATPase; strong similarity to peroxisome

biosynthesis protein PAB1 (PID:G1172019); coded for by

human CDNA C04279 (NID:G1465530)"

/codon_start=1

/protein_id="AAB46346.1"

/db_xref="GI:1669371"

/translation="KRLNIQKTELEVAFAVMNQPSVLLDDILTAGLPVPEHEH

SPDAORCEILCNVKNKIDCDINKFTDLDLOHAKKEGVAADFTLVDRALHSRL

SRQISIREKIVITLTDPOKALRGFRLPSLSVNLHKRDGCMKIGLHVRQILMD

TIQPAKVCLEKKEKPELFLANLPIRQRTGILLYGPFGTKLLAGVLAESRRNFISY

KLPBLISKYIGASQAVADIFIRQAAPCILFDEFESIAPRGHDTGTVDAVVKQ

LITQIDGVEGIGQVLAATSPDLIDPALRPGRLDCVCPDPDQVTTISLESKTO

OMLSFVSRLEIINVLSDSLPADVDVLOHVASVTSFTGADIKALINYNLEALHG

MLSKMSRIILPDESKFNMYRFGSSVSELSGNGTSSDLSQCLSAPESTMQDLPV

GKDDLFSGPYLRTASQEGCELTQEQQRDQRLADISITIKGKYSQSGEDSMQPGPI

KTRLAISQSHLMTALGHTRPISIEDWKNFAEL"

/complement(4948..5130)

/rpt_family="ALU"

/complement(6581..7133)

/rpt_family="L1"

/complement(7767..8037)

/rpt_family="ALU"

/complement(8186..8472)

/rpt_family="ALU"

8473..8625

/gene="WUGSC:H_RG083M05.1"

/note="match to human 3' EST H75782 (NID:G1049794), bases

287-444"

8841..9161

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H75921 (NID:G1050050), bases

21-348"

9481..9547

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST N22627 (NID:G1130501), bases

276-343"

repeat_region

complement(12612..12907)

misc_feature

/rpt_family="ALU"

13670..13793

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:G17434), bases

143-266"

repeat_region

13794..13877

misc_feature

/rpt_family="ALU"

13878..13906

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:G17434), bases

30-58"

repeat_region

13907..14104

repeat_region

/rpt_family="ALU"

complement(14110..14137)

repeat_region

/rpt_family="L1"

complement(15618..15907)

repeat_region

/rpt_family="ALU"

17227..17522

misc_feature

/rpt_family="ALU"

18667..19235

repeat_region

/note="match to human fetal brain 5' EST D61494

(NID:G970409), bases 1-255, and to human 3' EST R07476

(NID:G759399)"

repeat_region

19550..19670

misc_feature

/rpt_family="ALU"

21507..37303

/note="similarity to various SS-RNA virus polypeptides;
pseudogene; region of matches and close matches to
multiple human ESTs, see R68740 (NID:G842257)"

exon

37316..37489

repeat_region

/note="Grail prediction, score = 80"

complement(38938..39224)

repeat_region

/rpt_family="ALU"

39225..39707

misc_feature

/note="match to multiple human ESTs, see N30113

(NID:G1148633)"

repeat_region

39800..40085

repeat_region

/rpt_family="ALU"

complement(40247..40538)

repeat_region

/rpt_family="ALU"

complement(40632..40924)

repeat_region

/rpt_family="ALU"

complement(42283..42891)

repeat_region

/rpt_family="ALU"

complement(44574..45613)

repeat_region

/rpt_family="ALU"

complement(45614..45737)

misc_feature

/note="match to human 3' EST H48898 (NID:G988738), bases

129-333"

repeat_region

complement(46107..47026)

repeat_region

/note="match to multiple human ESTs, see N81064

(NID:G1243765), H48897 (NID:G988737), and W78831

(NID:G273146)"

repeat_region

complement(47027..47318)

repeat_region

/rpt_family="ALU"

complement(47365..47782)

repeat_region

/note="match to multiple human ESTs, see W37495

(NID:G1319089)"

repeat_region

47898..48115

misc_feature

/note="match to human 5' EST H62306 (NID:G1015138), bases

93-368"

repeat_region

complement(48116..48405)

repeat_region

/rpt_family="ALU"

complement(48406..48584)

repeat_region

/note="match to human 3' EST N29952 (NID:G1148472), bases

290-455, and 5' EST R12730 (NID:G765806)"

repeat_region

complement(48787..49405)

repeat_region

/rpt_family="ALU"

complement(49406..49534)

repeat_region

/note="match to human 3' EST R65794 (NID:G838432), bases

309-440"

```
repeat_region      complement(49638..45672)
                    /rpt_family="ALU"
misc_feature        complement(49674..49890)
                    /note="match to human 3', EST N29952 (NID:g1148472) and 5'
                    EST N29938 (NID:g1148458), sequences are from opposite
                    ends of the same clone"
gene                complement(49698..51806)
                    /gene="WUGSC:H.RG083M05.2"
                    complement(join(49698..49898,51575..51806))
                    /note="WUGSC:H.RG083M05.2"
                    /note="coded for by human cDNA K37389 (NID:g1319205),
                    R65891 (NID:g838529), R65794 (NID:g838432) and R65794
                    (NID:g838432)"
CDS                 /codon_start=1
                    /protein_id="AAB46345.1"
                    /db_xref="GI:1669370"
                    /translation="MFYFQVGIIFPCPGVYVIOGVVSVIBDQKPYVAQIRGR
                    IQDYCEKSAALTWLIFPLSSIRDFDPASTYIGEEEDLPKMEYLEFVCHAPSEYFK
                    SRSSPFVPTPRPEKGYIWHVGPPTPAITIKESVANHL"
exon                complement(51576..51758)
                    /gene="WUGSC:H.RG083M05.2"
                    /note="Grail prediction, score = 86"
                    /evidence="not experimental"
repeat_region      complement(52052..52329)
                    /rpt_family="L1"
misc_feature        complement(52052..52329)
                    /note="match to human EST M79192 (NID:g273505) base 2-289"

Query Match      100.0%; Score 501; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. No. 2,36-138;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CAAGATCTCAGATTATCAATGAGGCTGTTGTTCCCTATACGAGCTTACCTGACCCCT 60
DB 33000 CAGAGTCTGAGATTATCAATGAGGCTGTTGTTCCCTATACGAGCTTACCTGACCCCT 33059
QY 61 TATACTCTGCTTTCCCAATATACGAGAGAGCAAGTGTGTACAGTCTTGACCTTCAG 120
DB 33066 TATACTCTGCTTTCCCAATATACGAGAGAGCAAGTGTGTACAGTCTTGACCTTCAG 33119
QY 121 GATGCTTTTCTTGACATCCCTGTACATCTCTGACCTTCAATTTCTTGTTGCTTTGAAGT 180
DB 33120 GATGCTTTTCTTGACATCCCTGTACATCTCTGACCTTCAATTTCTTGTTGCTTTGAAGT 33179
QY 181 ACTTGAACCAACAATCTCAATCTGAGACTTTTATCCCAAGGCTTCAAGGAGTACT 240
DB 33180 ACTTGAACCAACAATCTCAATCTGAGACTTTTATCCCAAGGCTTCAAGGAGTACT 33239
QY 241 CCCCATCTATTGGCCAGCATTTAGCCCAAGACTTGAGCAATCTCATACCTGACACT 300
DB 33240 CCCCATCTATTGGCCAGCATTTAGCCCAAGACTTGAGCAATCTCATACCTGACACT 33299
QY 301 TGTCTTGGTAGTGATGATTTACTTTTGGCCGCCATTCAGAAACCTTGTCATCA 360
DB 33300 TGTCTTGGTAGTGATGATTTACTTTTGGCCGCCATTCAGAAACCTTGTCATCA 33359
QY 361 ACCCAACCAAGGCTTCAATTTCTGCTACCTGTGCTACATGCTTCCAAACCA 420
DB 33360 ACCCAACCAAGGCTTCAATTTCTGCTACCTGTGCTACATGCTTCCAAACCA 33419
QY 421 GGCTCAACTCTGCTACAGAGTTACTTAGGCTTAAATTAATTCAGAGCAACGAGGCC 480
DB 33420 GGCTCAACTCTGCTACAGAGTTACTTAGGCTTAAATTAATTCAGAGCAACGAGGCC 33479
QY 481 CTCAGTAGAGAACATCCAG 501
DB 33480 CTCAGTAGAGAACATCCAG 33500
```

```
RESULT 4
AC007566/c AC007566 149194 bp DNA linear PRI 01-MAR-2002
LOCUS
DEFINITION Homo sapiens BAC clone CTB-10G5 from Tg21-7q22, complete sequence.
```

ACCESSION AC007566
VERSION AC007566.2 GI:11181861
KEYWORDS HTG.
ORGANISM Homo sapiens.
SOURCE Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 149194)
AUTHORS Sulston, J.B. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 149194)
AUTHORS Du, Z.
TITLE The sequence of Homo sapiens BAC clone CTB-10G5
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 149194)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 149194)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 149194)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 149194)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 7 (bases 1 to 149194)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 8 (bases 1 to 149194)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 16, 2000 this sequence version replaced gi:4835815.
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_RG010G05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:regreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-9785K-B. The library contains cloned DNA from the male fibroblast cell line 9785K. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-91H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES Location/Qualifiers

source

```
1. 149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_lib="CTB-9785K-B"
1. 1634
/rpt_family="L1"
misc_feature
2248..2388
/note="match to EST BG752883 (NID:g14063536)"
misc_feature
2248..2387
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
2248..2387
/note="match to EST AW579261 (NID:g7254310)"
misc_feature
2248..2387
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
2248..2387
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
2248..2374
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
2248..2287
/note="match to EST AV686676 (NID:g10288539)"
2253..2387
/note="match to EST AU123510 (NID:g10948226)"
2344..2387
/note="match to EST AV686676 (NID:g10288539)"
2696..3066
/rpt_family="Alu"
3108..3392
/rpt_family="Alu"
misc_feature
3540..3628
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
3540..3628
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
3542..3632
/note="match to EST AV686676 (NID:g10288539)"
3542..3628
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
3542..3628
/note="match to EST AU123510 (NID:g10948226)"
3542..3628
/note="match to EST BG766882 (NID:g14077535)"
```

```
misc_feature 3542..3628
/note="match to EST BI160365 (NID:g14620366)"
misc_feature 3542..3609
/note="match to EST AW579261 (NID:g7254310)"
misc_feature 3562..3628
/note="match to EST BE272564 (NID:g9183456)"
misc_feature 3598..3628
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3714..3785
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 3717..3785
/note="similar to Mus musculus EST BE994936 (NID:g10678674)"
misc_feature 3719..3785
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 3719..3785
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature 3719..3785
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 3719..3785
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 3719..3785
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3719..3785
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 3719..3785
/note="match to EST BI160365 (NID:g14620366)"
misc_feature 3719..3785
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature 3860..3862
/note="match to EST AV686676 (NID:g10288539)"
repeat_region 4730..4827
/rpt_family="L2"
repeat_region 4834..4925
/rpt_family="L1"
misc_feature 5715..5860
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature 5715..5860
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 5715..5860
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 5715..5860
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 5715..5860
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 5715..5860
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 5715..5860
/note="match to EST BI160365 (NID:g14620366)"
misc_feature 5715..5860
/note="similar to Mus musculus EST BE994936 (NID:g10678674)"

Query Match 100.0%; Score 501; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 2..2e-138;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGGATTAATCATGAGGCTGTTCTCTATATGCCAGCTGTACTAGCCCT 60
DB 89004 CAAGATCTCAGGATTAATCAAGAGGCTGTTCTCTATATGCCAGCTGTACTAGCCCT 88945
QY 61 TATATCTGCTTTCCCAATATCCAGAGGAGAGGAGGTTTACAGTCTGAGCTTTCAG 120
DB 88944 TATATCTGCTTTCCCAATATCCAGAGGAGAGGAGGTTTACAGTCTGAGCTTTCAG 88885
QY 121 GATGCTTCTTCTGATCCCTGTAATCTGACTCTCAATCTTGTGCTTGAAGAT 180
DB 88884 GATGCTTCTTCTGATCCCTGTAATCTGACTCTCAATCTTGTGCTTGAAGAT 88825
QY 181 ACTTCAACCAACATCTCACTCACTGAGTATTTTACCCCAAGGTTTACGGATAGT 240
DB 88824 ACTTCAACCAACATCTCACTCACTGAGTATTTTACCCCAAGGTTTACGGATAGT 88765
```

QY 241 CCCCATTATTGGCCAGGCAATTAGCCCAAGATTGAGCCAACTCTCATPACCTGAGACT 300
Db 88764 CCCCATTATTGGCCAGGCAATTAGCCCAAGATTGAGCCAACTCTCATPACCTGAGACT 88705
QY 301 TGTCTTGGTGAAGTGAATGATTTCTTTGGGCCCCATTTCAGAAACCTTGTGGCACA 360
Db 88704 TGTCTTGGTGAAGTGAATGATTTCTTTGGGCCCCATTTCAGAAACCTTGTGGCACA 88645
QY 361 AGCCACCCAGGCGCTCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCAA 420
Db 88644 AGCCACCCAGGCGCTCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCAA 88585
QY 421 GGCTCAACTGTGCTACAGCAGGTTACTTAAAGCCTAAATATTCAGAAAGCAGGCGCC 480
Db 88584 GGCTCAACTGTGCTACAGCAGGTTACTTAAAGCCTAAATATTCAGAAAGCAGGCGCC 88525
QY 481 CTGAGTGAGGAACATCTCCAG 501
Db 88524 CTGAGTGAGGAACATCTCCAG 88504

RESULT 5
AX00957 2938 bp DNA linear PAT 10-MAR-2000
LOCUS Sequence 2 from Patent WO902695.
DEFINITION AX00957
ACCESSION AX00957.1 GI:7241199
VERSION
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
REFERENCE
1 (bases 1 to 2938)
Beseme, F. and Blond, J.
ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
Patent: WO 9902696-A 2 21-JAN-1999;
JOURNAL BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source
1..2938
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN

Query Match 94.6%; Score 474; DB 6; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 CAAGATCTCAGGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGCTGACCTAGCC 59
Db 2079 CAAGATCTCAGGATTATCAATGAGGCGCTGTCTTTATACCCAGCTGACCTAGCC 2138
QY 60 TTAATCTGCTTTTCCCAATATCCAGAGAGACAGAGTGGTTTACAGTCTTGACCTTCA 119
Db 2139 TTAATCTGCTTTTCCCAATATCCAGAGAGACAGAGTGGTTTACAGTCTTGACCTTCA 2198
QY 120 GGATGCTTCTTGTGATCCCTGTACATCTGACCTGACCTCAATCTTGTGTCCTTGAAGA 179
Db 2199 GGATGCTTCTTGTGATCCCTGTACATCTGACCTGACCTCAATCTTGTGTCCTTGAAGA 2258
QY 180 TACTTCAACCAACATCTCACTGACTATTTTACCCCAAGGTTGAGGATAG 239
Db 2259 TACTTCAACCAACATCTCACTGACTATTTTACCCCAAGGTTGAGGATAG 2318
QY 240 TCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCAACTCTCATPACCTGAGAC 299
Db 2319 TCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCAACTCTCATPACCTGAGAC 2378
QY 300 TGTCTTGGTGAAGTGAATGATTTCTTTGGGCCCCATTTCAGAAACCTTGTGGCACA 359
Db 2379 TGTCTTGGTGAAGTGAATGATTTCTTTGGGCCCCATTTCAGAAACCTTGTGGCACA 2438
QY 360 AAGCCACCAAGCGCTCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCAA 419
Db 2559 AAGCCACCAAGCGCTCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCAA 419

Db 2439 AAGCCACCAAGCGCTCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCAA 2498
QY 420 AGGCTCAACTGTGCTACAGCAGGTTACTTAAAGGCTAAATTTATCCAAAGCAGGCGC 479
Db 2499 AGGCTCAACTGTGCTACAGCAGGTTACTTAAAGGCTAAATTTATCCAAAGCAGGCGC 2558
QY 480 CCTCAGTGAGGAACATCTCCAG 501
Db 2559 CCTCAGTGAGGAACATCTCCAG 2580

RESULT 6
AX027471 2938 bp DNA linear PAT 16-SEP-2000
LOCUS Sequence 21 from Patent FR278784.
DEFINITION AX027471
ACCESSION AX027471
VERSION AX027471.1 GI:10188435
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2938)
Maillet, F., Voisset, C. and Paranhos, B.G.
Patent: FR 278784-A 21 28-JUL-2000;
JOURNAL BIO MERIEUX (FR)
FEATURES
source
1..2938
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN

Query Match 94.6%; Score 474; DB 6; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 CAAGATCTCAGGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGCTGACCTAGCC 59
Db 2079 CAAGATCTCAGGATTATCAATGAGGCGCTGTCTTTATACCCAGCTGACCTAGCC 2138
QY 60 TTAATCTGCTTTTCCCAATATCCAGAGAGACAGAGTGGTTTACAGTCTTGACCTTCA 119
Db 2139 TTAATCTGCTTTTCCCAATATCCAGAGAGACAGAGTGGTTTACAGTCTTGACCTTCA 2198
QY 120 GGATGCTTCTTGTGATCCCTGTACATCTGACCTGACCTCAATCTTGTGTCCTTGAAGA 179
Db 2199 GGATGCTTCTTGTGATCCCTGTACATCTGACCTGACCTCAATCTTGTGTCCTTGAAGA 2258
QY 180 TACTTCAACCAACATCTCACTGACTATTTTACCCCAAGGTTGAGGATAG 239
Db 2259 TACTTCAACCAACATCTCACTGACTATTTTACCCCAAGGTTGAGGATAG 2318
QY 240 TCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCAACTCTCATPACCTGAGAC 299
Db 2319 TCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCAACTCTCATPACCTGAGAC 2378
QY 300 TGTCTTGGTGAAGTGAATGATTTCTTTGGGCCCCATTTCAGAAACCTTGTGGCACA 359
Db 2379 TGTCTTGGTGAAGTGAATGATTTCTTTGGGCCCCATTTCAGAAACCTTGTGGCACA 2438
QY 360 AAGCCACCAAGCGCTCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCAA 419
Db 2499 AAGCCACCAAGCGCTCTTCAATTTCTGCTACCTGTGAGTACATGTTTCCAAACCAA 2498
QY 420 AGGCTCAACTGTGCTACAGCAGGTTACTTAAAGGCTAAATTTATCCAAAGCAGGCGC 479
Db 2499 AGGCTCAACTGTGCTACAGCAGGTTACTTAAAGGCTAAATTTATCCAAAGCAGGCGC 2558
QY 480 CCTCAGTGAGGAACATCTCCAG 501
Db 2559 CCTCAGTGAGGAACATCTCCAG 2580

RESULT 7	AF072499	2938 bp	mRNA	linear	PRI 10-FEB-1999
LOCUS	AF072499				
DEFINITION	AF072499				
ACCESSION	AF072499				
VERSION	AF072499.1				
KEYWORDS	GI:4262283				
SOURCE					
ORGANISM	Homo sapiens.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 2938)				
TITLE	Blond, J.L., Beseme, F., Duret, L., Bouton, O., Bedin, F., Perron, H., Mandrand, B. and Mallet, F.				
JOURNAL	Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family				
MEDLINE	J. Virol. 73 (2), 1175-1185 (1999)				
PUBMED	9982319				
REFERENCE	2 (bases 1 to 2938)				
AUTHORS	Blond, J.-L., Beseme, F. and Mallet, F.				
TITLE	Direct Submission				
JOURNAL	Submitted (19-JUN-1998) UMI03 CNRS-biomeleux, ENS Lyon, 46 allée d'Italie, Lyon, Cedex 07 69364, France				
FEATURES	Location/Qualifiers				
source	1..2938				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="cl.6A1"				
	/tissue_type="Placenta"				
	/clone_lib="Clontech 5'-stretch plus library, Cat number HL5014g"				
misc_feature	1..2938				
	/note="corresponds to gag and pol"				
repeat_region	1..2938				
	/rpt_family="HERV-W"				
BASE COUNT	878 a 720 c 646 g 692 t 2 others				
ORIGIN					
Query Match	94.6%; Score 474; DB 9; Length 2938;				
Best Local Similarity	97.8%; Pred. No. 3e-130;				
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;					
1	CAAGATCTCAGGATATCAAT-GAGGCTGTGTCTCTATGACGAGCTGTACCTAGGCC	59			
2079	CAAGATCTCAGGATATCAATGAGGAGGCGGTCTTATACCAAGCTGTACCTAGGCC	2138			
60	TTATATCTGTGCTTCCCAATATCAGAGAGACAGAGTGTATTAAGTCTGTGACCTTCA	119			
2139	TTATATCTGTGCTTCCCAATATCAGAGAGACAGAGTGTATTAAGTCTGTGACCTTCA	2198			
120	GGATGCTTCTTCTGTGATCCTGTATCATCTGACTCTCAATCTTGTGTGCTTTGAGA	179			
2199	GGATGCTTCTTCTGTGATCCTGTATCATCTGACTCTCAATCTTGTGTGCTTTGAGA	2258			
180	TACTTCAAAACCAATCTCAATCTCACTGAGCTATTTTAAACCCCAAGGTTAGGGATAG	239			
2259	TACTTCAAAACCAATCTCAATCTCACTGAGCTATTTTAAACCCCAAGGTTAGGGATAG	2318			
240	TCCCATCTATTTGGCGCAGGACATTAGCCCAAGTGTAGCAATCTCTACCTGTGACAC	299			
2319	CCCCATCTATTTGGCGCAGGACATTAGCCCAAGTGTAGCAATCTCTACCTGTGACAC	2378			
300	TTGTCTCTGTGAGTGTGATGATTTACTTTTGGCGGCCCATTCAGAAACCTTGTGCTATC	359			
2379	TTGTCTCTGTGAGTGTGATGATTTACTTTTGGCGGCCCATTCAGAAACCTTGTGCTATC	2438			
360	AAGCACCACCAAGCGCTTCAATTTCTCTGCACTGAGGCGGCTCAATGTTTCCAAACCA	419			
2439	AAGCACCACCAAGCGCTTCAATTTCTCTGCACTGAGGCGGCTCAATGTTTCCAAACCA	2498			

QY	420	AGGCTCAACTCGCTGCACAGAGGTTACTTGGGTAAATTTATCCAAAGGACACAGGAC	479
Db	2499	AGGCTCAACTCTGCTCTACAGAGGTTACTTGGGTAAATTTATCCAAAGGACACAGGAC	2558
QY	480	CCTCAGTAGAGNAACATCCAG 501	
Db	2559	CCTCAGTAGAGNAACATCCAG 2580	
RESULT 8			
LOCUS	AX000966	7582 bp	DNA linear PAT 10-MAR-2000
DEFINITION	Sequence 11 from Patent WO9902656.		
ACCESSION	AX000966		
VERSION	AX000966.1	GI:7241208	
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unidentified		
REFERENCE	1 (bases 1 to 7582)		
AUTHORS	Beesme, F. and Blond, J.		
TITLE	ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE DISEASES OR WITH PREGNANCY DISORDERS		
JOURNAL	Patent: WO 9902696-A 11 21-JAN-1999		
FEATURES	BIO MERIEUX (FR); BESEME FREDERIC (FR)		
source	Location/Qualifiers		
	1..7582		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	2156 a 1876 c 1538 g 1796 t	216 others	
ORIGIN			
Query Match	88.2%; Score 442; DB 6; Length 7582;		
Best Local Similarity	91.3%; Pred. No. 1e-120;		
Matches	464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;		
QY	1	CAGATCTCAGAGTATATAT-GAGCGTGTGTTCTCTATAGCCAGCTGTACTAGCCC	59
Db	2678	CAGAGTCTCAGAGTATATCAAGAGGCGGTGTCTTTATACCAAGCTGTACTAGCCC	2737
QY	60	TTATCTCGCTTCCCAATATACAGAGAGGAGAGTGTTTACAGTCCAGTCCAGACTTCA	119
Db	2738	TTATCTCTGTGTTTCCCAATATACAGAGAGGAGAGTGTTTACAGTCCAGTCCAGACTTCA	2797
QY	120	GGATGCTTCTTCTGCATCCCTGTATCATCTGCATCTCAATTCTTGTGCTTTGAAGA	179
Db	2798	GGATGCTTCTTCTGCATCCCTGTATCATCTGCATCTCAATTCTTGTGCTTTGAAGA	2857
QY	180	TACTTCAAAACCAACATCTCACTGACTGTATTTTACCCCAAGGTTTCAAGGATAG	239
Db	2858	TACTTCAAAACCAACATCTCACTGACTGTATTTTACCCCAAGGTTTCAAGGATAG	2917
QY	240	TCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCAATCCATACCTGGACA-	298
Db	2918	TCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCAATCCATACCTGGACA-	2977
QY	299	CTTGTCTCTGGTATGATGATTTACTTTTGGCCGCCATCAGAGAACTTTGGCA	357
Db	2978	CTTGTCTCTGGTATGATGATTTACTTTTGGCCGCCATCAGAGAACTTTGGCA	3037
QY	358	TCAAGCCACCAAGGCTTCAATTCTGCTACCTGTGGCTACATGTTTCCAAAC	417
Db	3038	TCAAGCCACCAAGGCTTCAATTCTGCTACCTGTGGCTACATGTTTCCAAAC	3097
QY	418	AAAGCTCAACTCTGCTACAGCAGT---TACTTAGGCTTAAATTTATCCAAAGGAC	473
Db	3098	AAAGCTCAACTCTGCTACAGCAGT---TACTTAGGCTTAAATTTATCCAAAGGAC	3157
QY	474	CAGGCGCTCAGTAGGAACATCCAG 501	
Db	3158	CAGGCGCTCAGTAGGAATTCAG 3185	

LOCUS	AX027480	7582 bp	DNA	linear	PAT 16-SEP-2000
DEFINITION	Sequence 30 from Patent FR2788794.				
ACCESSION	AX027480				
VERSION	AX027480.1	GI:10188444			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 7582)				
JOURNAL	Mallet, F., Voisset, C. and Paranhos, B. G. Patent: FR 2788784-A 30 28-JUL-2000; BIO MERIEUX (FR)				
FEATURES	Location/Qualifiers				
source	1..7582				
BASE COUNT	2156 a 1876 c 1538 g 1796 t 216 others				
ORIGIN	1..7582				
Query Match	88.2%; Score 442, DB 6; Length 7582;				
Best Local Similarity	91.3%; Pred. No. 1e-120;				
Matches	464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;				
QY	1 CAAGATCTCAGATTATCAAT-GAGGCTGTGTCTCTATATGACCAAGCTGTAACCC 59				
DB	2678 CAAGATCTCAGATTATCAATGAGGCGCTGTCTTATATACCAAGCTGTAACCC 2737				
QY	60 TTATATCTGCTTCCCAATACGAGGAACG/GAGTGGTTACAGCTCGAGACTTCA 119				
DB	2738 TTATATCTGCTTCCCAATACGAGGAACG/GAGTGGTTACAGCTCGAGACTTCA 2797				
QY	120 GGATCGCTTCTCTGCACTCCGTACATCTCAATCTCAATCTGTTGCTTTGAAGA 179				
DB	2798 GGATCGCTTCTCTGCACTCCGTACATCTCAATCTCAATCTGTTGCTTTGAAGA 2857				
QY	180 TACTTCAAAACCAACATCTCAACTCACTGACIATTTTACCACCAAGGTTCAAGGATG 239				
DB	2858 TACTTCAAAACCAACATCTCAACTCACTGACIATTTTACCACCAAGGTTCAAGGATG 2917				
QY	240 TCCCATCTAATTGGCCAGGCAATTAGCCCAACACTGAGCCCAATCCTCACTGAGAC- 298				
DB	2918 TCCCATCTAATTGGCCAGGCAATTAGCCCAACACTGAGACIATTTTACCACCAAGGTTCAAGGATG 2977				
QY	299 -CTTGCTCTCGGTAGTGGATGATTTACTTTTGGCGCCCAATTCAGAAACCTTGCCA 357				
DB	2978 TCTTGCTCTCGGTAGTGGATGATTTACTTTTGGCGCCCAATTCAGAAACCTTGCCA 3037				
QY	358 TCAAGCACCACCAAGCGCTTCAATTTCTCGTACTCTGCTAGCTGAGCTCATGTTCCAAAC 417				
DB	3038 TCAAGCACCACCAAGCGCTTCAATTTCTCGTACTCTGCTAGCTGAGCTCATGTTCCAAAC 3097				
QY	418 AAAAGTCAACTCTGCTACACAGAGT----TACTTAAAGGCTTAAATTTCCAAAGGCAC 473				
DB	3098 AAAAGTCAACTCTGCTACACAGAGT----TACTTAAAGGCTTAAATTTCCAAAGGCAC 3157				
QY	474 CAGGCGCTCAGTGAGGACACATCCAG 501				
DB	3158 CAGGCGCTCAGTGAGGACACATCCAG 3185				
RESULT 10	AF045450/c	40205 bp	DNA	linear	PRI 20-MAR-1998
LOCUS	AF045450				
DEFINITION	Homo sapiens chromosome 21q22.3; cosmid Q11M15, complete sequence.				
ACCESSION	AF045450				
VERSION	AF045450.1	GI:2895783			
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				

REFERENCE	REFERENCE
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (bases 1 to 40205)
Taudien, S. and Rosenthal, A.	Unpublished
2 (bases 1 to 40205)	
Taudien, S., Nordstiek, G., Dagand, E., Hildmann, T., Drescher, B., Weber, J., Rosenthal, A. and Yaspo, M. L.	Direct Submission
Submitted (29-JUN-1998) Genome Analysis, Institut for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany	
Location/Qualifiers	
1. 40205	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="21"	
/map="21q22.3"	
/clone="cosmid Q11M15"	
/complement (293. 586)	
/evidence=not experimental	
/rpt_family="AluSc"	
/complement (643. 966)	
/evidence=not experimental	
/rpt_family="MERVL"	
/complement (1053. 1176)	
/note="Genscan, score = 5.19%, comment = Internal_exon 124 bp frame: 2 phase: 1"	
/evidence=not experimental	
/complement (1251. 1859)	
/evidence=not experimental	
/rpt_family="HERVL"	
/complement (2296. 3320)	
/evidence=not experimental	
/rpt_family="HERVL6"	
2818. 2889	
/note="Xpound exon prediction, score = 82% (0%)"	
/evidence=not experimental	
/complement (321. 3658)	
/evidence=not experimental	
/rpt_family="MLT1C"	
3359. 3395	
/note="Xpound exon prediction, score = 84% (0%)"	
/evidence=not experimental	
/complement (3874. 4176)	
/evidence=not experimental	
/rpt_family="Aluub"	
/complement (4330. 4433)	
/evidence=not experimental	
/rpt_family="L2"	
4585. 4638	
/note="GRAIL, score = 95.000%, comment = excellent shadow"	
/evidence=not experimental	
/complement (5113. 7899)	
/evidence=not experimental	
/rpt_family="L1PA14"	
/complement (6119. 6598)	
/note="GRAIL, score = 63.000%, comment = good"	
/evidence=not experimental	
/complement (7977. 8045)	
/evidence=not experimental	
/rpt_family="FLAM_A"	
8048. 8178	
/evidence=not experimental	
/rpt_family="L2"	
8771. 8876	
/note="MZEF, score = 67.4%"	
/evidence=not experimental	
/complement (8809. 9187)	
/evidence=not experimental	
/rpt_family="MERIA"	
8837. 8876	
/note="GRAIL, score = 55.000%, comment = good shadow"	
/evidence=not experimental	
/complement (9113. 9213)	

/note="GRAIL, score = 55.000%, comment = good"
/evidence=not_experimental
repeat_region complement(9130. .9258)
/evidence=not_experimental
/rpt_family="L1MC3"
9259. .9647
/evidence=not_experimental
/rpt_family="L1MC3"
repeat_region complement(10138. .10427)
/evidence=not_experimental
/rpt_family="MLT2E"
10662. .10875
/evidence=not_experimental
/rpt_family="MER58A"
exon complement(11072. .11156)
/note="GRAIL, score = 86.000%, comment = excellent"
/evidence=not_experimental
exon complement(11072. .11129)
/note="Xpound exon prediction, score = 89% (0%)"
/evidence=not_experimental
repeat_region 11478. .11597
/evidence=not_experimental
/rpt_family="L2"
11570. .11704
/note="WZEF, score = 66.1%"
/evidence=not_experimental
11570. .11704
/note="GRAIL, score = 99.000%, comment = excellent shadow"
/evidence=not_experimental
repeat_region complement(11687. .11735)
/evidence=not_experimental
/rpt_family="MIR"
repeat_region complement(11774. .12320)
/evidence=not_experimental
/rpt_family="MLT1B"
exon complement(11828. .11949)
/note="GRAIL, score = 50.000%, comment = good"
/evidence=not_experimental
12623. .12843
/evidence=not_experimental
/rpt_family="AlusX"
repeat_region 13102. .13457
/evidence=not_experimental
/rpt_family="THE1B"
exon complement(13481. .13519)
/note="GRAIL, score = 65.000%, comment = good"
/evidence=not_experimental
15065. .15188
/evidence=not_experimental
/rpt_family="HERVL"
misc_feature 15130. .15378
/note="GC score = 10.60 (249bp); Region: GC content"
/evidence=not_experimental
repeat_region 15369. .16268
/evidence=not_experimental
/rpt_family="MERVL"
exon complement(16623. .16714)
/note="GRAIL, score = 72.000%, comment = good"
/evidence=not_experimental
16661. .16730
/note="WZEF, score = 93.5%"
/evidence=not_experimental
16880. .17318
/evidence=not_experimental
/rpt_family="MLT1C"
repeat_region 17326. .17436
/evidence=not_experimental
/rpt_family="LTR16C"
repeat_region complement(17540. .17675)
/evidence=not_experimental
/rpt_family="MIR"
repeat_region complement(17805. .17864)
/evidence=not_experimental

exon /rpt_family="MLT1C"
18076. .18098
/note="GRAIL, score = 45.000%, comment = marginal shadow"
/evidence=not_experimental
exon complement(18219. .18349)
/note="Genscan, score = 2.99%, comment = Internal_exon 131
bp frame: 1 phase: 2"
/evidence=not_experimental
repeat_region complement(18717. .19211)
/evidence=not_experimental
/rpt_family="MLT1C"
exon complement(19252. .19399)
/note="GRAIL, score = 71.000%, comment = good"
/evidence=not_experimental
exon 19269. .19350
/note="GRAIL, score = 81.000%, comment = excellent shadow"
/evidence=not_experimental
exon complement(19379. .19399)
/note="Xpound exon prediction, score = 60% (0%)"
/evidence=not_experimental
exon complement(19501. .19881)
/note="WZEF, score = 50.6%"
/evidence=not_experimental
repeat_region 19945. .20020
/note="homology = 100.00%, score = 38, counts = 2"
/evidence=not_experimental
/rpt_type=tandem
/rpt_unit=tctctgacactctctgtgagaggtcggttatcgc
repeat_region 20223. .20643
/evidence=not_experimental
/rpt_family="MLT1B"
repeat_region 20645. .20910
/evidence=not_experimental
/rpt_family="MLT1F"
repeat_region 20961. .21007
/evidence=not_experimental
/rpt_family="MLT1F"
repeat_region 21483. .21781
/evidence=not_experimental
/rpt_family="AluDb"
exon complement(21504. .21756)
/note="WZEF, score = 79.4%"
/evidence=not_experimental
repeat_region complement(22325. .22806)
/evidence=not_experimental
/rpt_family="MLT1C"
Query Match 86.8%; Score 435; DB 9; Length 40205;
Best Local Similarity 93.9%; Pred. No. 1,1e-118;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
QY 1 CAAGATCTCAGAGATTATCAATGAGGCTGTTTCTCTATATGCCAGCTGTACTAGCCCT 60
Db 30626 CAAGATCTCAGAGATTATCAATGAGGCTGTTTCTCTATATGCCAGCTGTACTAGCCCT 30567
QY 61 TATCTCGCTTCCCAATATCCAGAGAGACAGAGGTTTACAGTCTGAGACTTTCAG 120
Db 30566 TATCTCGCTTCCCAATATCCAGAGAGACAGAGGTTTACAGTCTGAGACTTTCAG 30507
QY 121 GATGCTTCTTCTGATCCCTGTATCATCTGACTCTCAATCTTGTGGCCTTGAAGAT 180
Db 30506 GATGCTTCTTCTGATCCCTGTATCATCTGACTCTCAATCTTGTGTGACCTTGAAGAT 30447
QY 181 ACTTCAACCCCAATCTCACTCACTGTGACGTATTTTATCCCAAGGTTTACGGATAGT 240
Db 30446 CCTTCAACCCCAATCTCACTCACTGTGACGTATTTTATCCCAAGGTTTACGGATAGT 30387
QY 241 CCCCATGATTGGCGAGGAGCTTAGCCCAAGACTGAGCCCAATCTCATCTGAGACT 300
Db 30386 CCCCATGATTGGCGAGGAGCTTAGCCCAAGACTGAGCCCAATCTCATCTGAGACT 30327
QY 301 --TGTCTTGGTGGATGATGATTTACTTTGGCCGCCCAATTGAAACCTTGTGCAT 358
|||||

Db	30326	CCTGCTCTTACGTACGATGATGATTTACTTTTACGTCCCGCTTCAGAAACCTTTGGCCAT	30267	repeat_region	2629. .2831	/evidence=not experimental	
Qy	359	CAAGCCACCCAAAGGCTCTTCAATTTCTCGTCACTGCTGGCTACATGGTTTCCAAACA	418	/rpt_family="AluXs"			
Db	30266	CAAGCACCACCAAGGCTCTTAAATTTCTCACTACCTGTGGCTACAAAGTTTCCAAACA	30207	repeat_region	3012. .3122	/evidence=not experimental	
Qy	419	AAAGCTCAACTCTGCTCAGACGAGGTTA----	474	/rpt_family="MLT1L"	3123. .3478	/evidence=not experimental	
Db	30206	AAAGCTCAGCTCTCTCAACAGCAGTTAAATGCTTAGGGCTAAATATTCACAAAGTCACC	30147	repeat_region	3479. .3618	/evidence=not experimental	
Qy	475	AGGGCCCTCAGTGGAGGAACACATCCAG	501	/rpt_family="THERB"			
Db	30146	AGGGCCCTCAGTGGAGGAACATTCACAG	30120	exon	4674. .4781	/evidence=not experimental	
RESULT 11				exon	4674. .4781	/note="GRAIL, score = 42.000%, comment = marginal"	
LOCUS	AF121782	142742 bp	DNA	linear	4712. .4749	/evidence=not experimental	
DEFINITION	Homo sapiens chromosome 21q22.3	PAC 206A10	complete sequence.	exon	4712. .4749	/note="Xpound exon prediction, score = 70% (0%)"	
ACCESSION	AF121782			repeat_region	5385. .6104	/evidence=not experimental	
VERSION	AF121782.1	GI:4210991		exon	6682. .6751	/rpt_family="HERVL"	
KEYWORDS	HTG.			exon	6682. .6751	/note="MZEP, score = 95.1%"	
SOURCE	Homo sapiens.			exon	6746. .6751	/evidence=not experimental	
ORGANISM	Homo sapiens.			exon	6746. .6751	/note="MZEP, score = 95.1%"	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Citerhini; Homnidae; Homo. 1 (bases 1 to 142742)			exon	6746. .6751	/note="Genscan, score = 1.83%, comment = initial_exon 6 bp frame: 1 phase: 0"	
AUTHORS	Taddei, S., Dagnid, E., Hildmann, T., Nordstiek, G., Drescher, B., Schatevov, R., Weber, J., Schilling, M., Menzel, U., Yaspo, M.L. and Rosenthal, A.			repeat_region	6901. .7339	/evidence=not experimental	
TITLE	Direct Submission			repeat_region	7347. .7457	/evidence=not experimental	
JOURNAL	Submitted (21-JAN-1999) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany			repeat_region	7347. .7457	/rpt_family="MLT1C"	
FEATURES	location/Dualifiers			repeat_region	7347. .7457	/evidence=not experimental	
source	1. 142742			repeat_region	7347. .7457	/rpt_family="Lrr16C"	
	/organism="Homo sapiens;"			repeat_region	7347. .7457	/evidence=not experimental	
	/db_xref="taxon:9606"			repeat_region	7347. .7457	/rpt_family="Lrr16C"	
	/chromosome="21"			repeat_region	7347. .7457	/evidence=not experimental	
	/map="21q22.3"			repeat_region	7347. .7457	/rpt_family="Lrr16C"	
	/clone="PAC 206A10"			repeat_region	7347. .7457	/evidence=not experimental	
	3. .118			repeat_region	7347. .7457	/rpt_family="MLT1L"	
	/evidence=not experimental			exon	8097. .8119	/note="GRAIL, score = 45.000%, comment = marginal"	
	/rpt_family="MIR"			repeat_region	8097. .8119	/evidence=not experimental	
	complement(140. .415)			repeat_region	8097. .8119	/evidence=not experimental	
	/evidence=not experimental			repeat_region	8097. .8119	/evidence=not experimental	
	/rpt_family="MLT2E"			repeat_region	8097. .8119	/evidence=not experimental	
	complement(668. .881)			repeat_region	8097. .8119	/evidence=not experimental	
	/evidence=not experimental			exon	8097. .8119	/rpt_family="MLT1C"	
	/rpt_family="HER58A"			exon	8097. .8119	/evidence=not experimental	
	complement(1078. .1162)			exon	8097. .8119	/evidence=not experimental	
	/note="GRAIL, score = 86.000%, comment = excellent shadow"			exon	8097. .8119	/note="GRAIL, score = 71.000%, comment = good"	
	/evidence=not experimental			exon	8097. .8119	/evidence=not experimental	
	1114. .1403			exon	8097. .8119	/note="GRAIL, score = 81.000%, comment = excellent shadow"	
	/evidence=not experimental			exon	8097. .8119	/evidence=not experimental	
	/rpt_family="MLT1L"			exon	8097. .8119	/evidence=not experimental	
	1481. .1603			exon	8097. .8119	/evidence=not experimental	
	/evidence=not experimental			exon	8097. .8119	/evidence=not experimental	
	/rpt_family="T2"			exon	8097. .8119	/evidence=not experimental	
	1576. .1710			exon	8097. .8119	/evidence=not experimental	
	/note="MZEP, score = 66.1%"			exon	8097. .8119	/evidence=not experimental	
	/evidence=not experimental			exon	8097. .8119	/evidence=not experimental	
	1576. .1710			exon	8097. .8119	/evidence=not experimental	
	/note="GRAIL, score = 99.000%, comment = excellent"			exon	8097. .8119	/evidence=not experimental	
	/evidence=not experimental			exon	8097. .8119	/evidence=not experimental	
	complement(1683. .1741)			exon	8097. .8119	/evidence=not experimental	
	/evidence=not experimental			exon	8097. .8119	/evidence=not experimental	
	/rpt_family="MIR"			exon	8097. .8119	/evidence=not experimental	
	complement(1780. .2326)			exon	8097. .8119	/evidence=not experimental	
	/evidence=not experimental			exon	8097. .8119	/evidence=not experimental	
	/						

/note="WZEF, score = 79.4%"
/evidence=not_experimental
complement(12346, .12827)
/evidence=not_experimental
/rpt_family="MULTIC"
12902, .13180
/evidence=not_experimental
/rpt_family="AluSg"
complement(13057, .13161)
/note="WZEF, score = 76.1%"
/evidence=not_experimental
complement(13326, .13364)
/evidence=not_experimental
/rpt_family="L1MC2"
complement(13613, .13908)
/evidence=not_experimental
/rpt_family="AluSx"
complement(14317, .14614)
/evidence=not_experimental
/rpt_family="LTR37A"
complement(14962, .15036)
/note="WZEF, score = 93%"
/evidence=not_experimental
15255, .15299
/evidence=not_experimental
/rpt_family="L1PA6"
complement(15469, .15497)
/note="Xpound exon prediction, score = 62% (0%)"
exon
repeat_region
/evidence=not_experimental
16045, .16768
/evidence=not_experimental
/rpt_family="MER21A"
complement(16555, .16655)
/note="GRAL, score = 54.000%, comment = good"
/evidence=not_experimental
16823, .16850
/evidence=not_experimental
/rpt_family="MER34"
complement(16851, .17208)
/evidence=not_experimental
/rpt_family="THE1A"
17209, .17632
/evidence=not_experimental
/rpt_family="MER34"
complement(17633, .17969)
/evidence=not_experimental
/rpt_family="AluSg1"
complement(18004, .18334)
/evidence=not_experimental
/rpt_family="LTR17"
complement(18335, .18632)
/evidence=not_experimental
/rpt_family="HERV17"
18332, .18420
/note="GRAL, score = 51.000%, comment = good shadow"
/evidence=not_experimental
complement(18397, .18584)
/note="GRAL, score = 66.000%, comment = good"
/evidence=not_experimental
complement(18643, .24647)
/evidence=not_experimental
/rpt_family="HERV17"
complement(19228, .19361)
/note="WZEF, score = 92.8%"
/evidence=not_experimental
20900, .21135
/note="WZEF, score = 51.4%"
/evidence=not_experimental
21150, .21275
/note="Xpound exon prediction, score = 74% (0%)"
/evidence=not_experimental
86.8%; Score 435; DB 9; Length 142742;

Best Local Similarity 93.9%; Pred. No. 1e-118;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
Qy 1 CAAGATCCAGAGTATTCATCAATGAGGCTGTGTCCTCATATGCGAGGCTGATCCAGCCCT 60
Db 20632 CAAGATCCAGAGTATTCATCAATGAGGCTGTGTCCTCATATGCGAGGCTGATCCAGCCCT 20573
Qy 61 TATATCTGCTTTCCCAATATCCAGAGAAACAGAGTGTATTACAGTCTTGACCTTCAG 120
Db 20572 TATATCTGCTTTCCCAATATCCAGAGAAACAGAGTGTATTACAGTCTTGACCTTCAG 20513
Qy 121 GATGCTTTCTTCTGCATCCCTGTAACATCTGACTCTCAATTCTGTTGCTTTGAAGAT 180
Db 20512 GATGCTTTCTTCTGCATCCCTGTAACATCTGACTCTCAATTCTGTTGCTTTGAAGAT 20453
Qy 181 ACTTCAAAACCAATCTCAATCTCACTGAGCTATTTTACCCCAAGGTTCAAGGATATGT 240
Db 20452 CTTTCAAAACCAATCTCAATCTCACTGAGCTATTTTACCCCAAGGTTCAAGGATATGT 20393
Qy 241 CCCCATCTATTGGCCAGGACATTAGCCCAAGACTTGAGCCATCTGATACCTTGACACT 300
Db 20392 CCCCATCTATTGGCCAGGACATTAGCCCAAGACTTGAGCCATCTGATACCTTGACACT 20333
Qy 301 --TGTCTTGGTAGTGTGATGATTTACTTTGGCCGCCATTCAGAAACCTTGCCAT 358
Db 20332 CTTGCTTTCAGTGCATGATGATTTACTTTAGCTGCCGCTGAGAAACCTTGCCAT 20273
Qy 359 CAAGCCACCCAGAGGCTCTTAAATTCCTGCTACCTGCTACATGCTTCCAAACCA 418
Db 20272 CAAGCCACCCAGAGGCTCTTAAATTCCTGCTACCTGCTACATGCTTCCAAACCA 20213
Qy 419 AAGGCTCACTCTGCTCAGACAGAGTTA---CTTAGGGCTAAATTTATCCAAAGGACAC 474
Db 20212 AAGGCTCACTCTGCTCAGACAGAGTTAATGCTTAGGGCTAAATTTATCCAAAGTAC 20153
Qy 475 AGGGCCTCAGTGAAGAACATCTCAG 501
Db 20152 AGGGCCTCAGTGAAGAACATCTCAG 20126
RESULT 12
HS21C080/c 34000 bp DNA linear PRI 24-MAY-2000
LOCUS
DEFINITION Homo sapiens chromosome 21 segment HS21C080.
ACCESSION AL163280 AP001735 BA000005
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 34000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Tokoi,Y., Choi,D.K., Seoda,E.,
Okui,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Aakawa,S., Shintani,A., Sasaki,T., Nagamine,K., MitsuYama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordtiek,G.,
Hornischer,K., Brandt,P., Scharfe,M., Scheen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Rieseemann,L., Dagand,E., Mehmer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaspo,M.L.
Direct Submission
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of

```
FEATURES
source
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
  Sagamihara 228-8555, Japan,
  * e.mail: sakaki@gsc.riken.go.jp
  * URL: http://hgp.gsc.riken.go.jp/
  and
  * Institute of Molecular Biotechnology, Genome Analysis, *
    Beutenbergstrasse 11, D-07745 Jena, Germany,
    * e.mail: gscj-submit@genome.imb-jena.de
    * URL: http://genome.imb-jena.de/
  and
  * Keio University School of Medicine, Dept. of Molecular Biology, *
    Tokyo 160-8582, Japan,
    * e.mail: shimizu@cmb.med.keio.ac.jp
    * URL: http://adenine.cmb.med.keio.ac.jp/
  and
  * GBF, Dept. of Genome Analysis,
    Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
    info.genome@gbf.de
    * URL: http://genome.gbf.de/
  and
  * Max-Planck Institute for Molecular Genetics,
    Inhestrasse 73, D-14195 Berlin, Germany,
    * e.mail: info-chr21@molgen.mpg.de
    * URL: http://chr21.rz-berlin.mpg.de/
    Location/Qualifiers
      1..340000
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="21"
        /map="21q22.3"
        <1..125946
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /chromosome="21"
          /map="21q22.3"
          /clone="P160L9, 5' partial"
          /clone_lib="RPC11,3-5 PAC library"
          /note="Accession No. AF121897"
          73410..243533
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="21"
            /map="21q22.3"
            /clone="P70124"
            /clone_lib="RPC11,3-5 PAC library"
            /note="Accession No. AF121782"
            224137..>340000
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="21"
              /map="21q22.3"
              /clone="P206A10, 3' partial"
              /clone_lib="RPC11,3-5 PAC library"
              /note="Accession No. AF121782"
              286628..>340000
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="21"
                /map="21q22.3"
                /clone="BAC-291B3, 3' partial"
                /clone_lib="BAC library"
                /note="Accession No. AF064857"
                1088..1566
                  /note="L1M4"
                  /rpt_family="LINE/L1"
                  /rpt_type=DISPERSED
                  2398..2495
                    /note="(TCTA)n"
                    /rpt_family="Simple_repeat"
                    /rpt_type=TANDEM
                    complement(2496..2878)
                    /note="THE1C"
                    /rpt_family="LTR/MaLR"

repeat_region
/rtc_type=DISPERSED
2879..2931
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
2932..3004
/note="(TCTA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3005..3074
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
3075..3157
/note="(SH3BGR)"
/rpt_type=TANDEM
3158..3195
/note="(SH3BGR)"
/rpt_type=TANDEM
3196..3233
/note="(SH3BGR)"
/rpt_type=TANDEM
3234..3271
/note="(SH3BGR)"
/rpt_type=TANDEM
3272..3309
/note="(SH3BGR)"
/rpt_type=TANDEM
3310..3347
/note="(SH3BGR)"
/rpt_type=TANDEM
3348..3385
/note="(SH3BGR)"
/rpt_type=TANDEM
3386..3423
/note="(SH3BGR)"
/rpt_type=TANDEM
3424..3461
/note="(SH3BGR)"
/rpt_type=TANDEM
3462..3499
/note="(SH3BGR)"
/rpt_type=TANDEM
3500..3537
/note="(SH3BGR)"
/rpt_type=TANDEM
3538..3575
/note="(SH3BGR)"
/rpt_type=TANDEM
3576..3613
/note="(SH3BGR)"
/rpt_type=TANDEM
3614..3651
/note="(SH3BGR)"
/rpt_type=TANDEM
3652..3689
/note="(SH3BGR)"
/rpt_type=TANDEM
3690..3727
/note="(SH3BGR)"
/rpt_type=TANDEM
3728..3765
/note="(SH3BGR)"
/rpt_type=TANDEM
3766..3803
/note="(SH3BGR)"
/rpt_type=TANDEM
3804..3841
/note="(SH3BGR)"
/rpt_type=TANDEM
3842..3879
/note="(SH3BGR)"
/rpt_type=TANDEM
3880..3917
/note="(SH3BGR)"
/rpt_type=TANDEM
3918..3955
/note="(SH3BGR)"
/rpt_type=TANDEM
3956..3993
/note="(SH3BGR)"
/rpt_type=TANDEM
3994..4031
/note="(SH3BGR)"
/rpt_type=TANDEM
4032..4069
/note="(SH3BGR)"
/rpt_type=TANDEM
4070..4107
/note="(SH3BGR)"
/rpt_type=TANDEM
4108..4145
/note="(SH3BGR)"
/rpt_type=TANDEM
4146..4183
/note="(SH3BGR)"
/rpt_type=TANDEM
4184..4221
/note="(SH3BGR)"
/rpt_type=TANDEM
4222..4259
/note="(SH3BGR)"
/rpt_type=TANDEM
4260..4297
/note="(SH3BGR)"
/rpt_type=TANDEM
4298..4335
/note="(SH3BGR)"
/rpt_type=TANDEM
4336..4373
/note="(SH3BGR)"
/rpt_type=TANDEM
4374..4411
/note="(SH3BGR)"
/rpt_type=TANDEM
4412..4449
/note="(SH3BGR)"
/rpt_type=TANDEM
4450..4487
/note="(SH3BGR)"
/rpt_type=TANDEM
4488..4525
/note="(SH3BGR)"
/rpt_type=TANDEM
4526..4563
/note="(SH3BGR)"
/rpt_type=TANDEM
4564..4601
/note="(SH3BGR)"
/rpt_type=TANDEM
4602..4639
/note="(SH3BGR)"
/rpt_type=TANDEM
4640..4677
/note="(SH3BGR)"
/rpt_type=TANDEM
4678..4715
/note="(SH3BGR)"
/rpt_type=TANDEM
4716..4753
/note="(SH3BGR)"
/rpt_type=TANDEM
4754..4791
/note="(SH3BGR)"
/rpt_type=TANDEM
4792..4829
/note="(SH3BGR)"
/rpt_type=TANDEM
4830..4867
/note="(SH3BGR)"
/rpt_type=TANDEM
4868..4905
/note="(SH3BGR)"
/rpt_type=TANDEM
4906..4943
/note="(SH3BGR)"
/rpt_type=TANDEM
4944..4981
/note="(SH3BGR)"
/rpt_type=TANDEM
4982..5019
/note="(SH3BGR)"
/rpt_type=TANDEM
5020..5057
/note="(SH3BGR)"
/rpt_type=TANDEM
5058..5095
/note="(SH3BGR)"
/rpt_type=TANDEM
5096..5133
/note="(SH3BGR)"
/rpt_type=TANDEM
5134..5171
/note="(SH3BGR)"
/rpt_type=TANDEM
5172..5209
/note="(SH3BGR)"
/rpt_type=TANDEM
5210..5247
/note="(SH3BGR)"
/rpt_type=TANDEM
5248..5285
/note="(SH3BGR)"
/rpt_type=TANDEM
5286..5323
/note="(SH3BGR)"
/rpt_type=TANDEM
5324..5361
/note="(SH3BGR)"
/rpt_type=TANDEM
5362..5399
/note="(SH3BGR)"
/rpt_type=TANDEM
5400..5437
/note="(SH3BGR)"
/rpt_type=TANDEM
5438..5475
/note="(SH3BGR)"
/rpt_type=TANDEM
5476..5513
/note="(SH3BGR)"
/rpt_type=TANDEM
5514..5551
/note="(SH3BGR)"
/rpt_type=TANDEM
5552..5589
/note="(SH3BGR)"
/rpt_type=TANDEM
5590..5627
/note="(SH3BGR)"
/rpt_type=TANDEM
5628..5665
/note="(SH3BGR)"
/rpt_type=TANDEM
5666..5703
/note="(SH3BGR)"
/rpt_type=TANDEM
5704..5741
/note="(SH3BGR)"
/rpt_type=TANDEM
5742..5779
/note="(SH3BGR)"
/rpt_type=TANDEM
5780..5817
/note="(SH3BGR)"
/rpt_type=TANDEM
5818..5855
/note="(SH3BGR)"
/rpt_type=TANDEM
5856..5893
/note="(SH3BGR)"
/rpt_type=TANDEM
5894..5931
/note="(SH3BGR)"
/rpt_type=TANDEM
5932..5969
/note="(SH3BGR)"
/rpt_type=TANDEM
5970..6007
/note="(SH3BGR)"
/rpt_type=TANDEM
6008..6045
/note="(SH3BGR)"
/rpt_type=TANDEM
6046..6083
/note="(SH3BGR)"
/rpt_type=TANDEM
6084..6121
/note="(SH3BGR)"
/rpt_type=TANDEM
6122..6159
/note="(SH3BGR)"
/rpt_type=TANDEM
6160..6197
/note="(SH3BGR)"
/rpt_type=TANDEM
6198..6235
/note="(SH3BGR)"
/rpt_type=TANDEM
6236..6273
/note="(SH3BGR)"
/rpt_type=TANDEM
6274..6311
/note="(SH3BGR)"
/rpt_type=TANDEM
6312..6349
/note="(SH3BGR)"
/rpt_type=TANDEM
6350..6387
/note="(SH3BGR)"
/rpt_type=TANDEM
6388..6425
/note="(SH3BGR)"
/rpt_type=TANDEM
6426..6463
/note="(SH3BGR)"
/rpt_type=TANDEM
6464..6501
/note="(SH3BGR)"
/rpt_type=TANDEM
6502..6539
/note="(SH3BGR)"
/rpt_type=TANDEM
6540..6577
/note="(SH3BGR)"
/rpt_type=TANDEM
6578..6615
/note="(SH3BGR)"
/rpt_type=TANDEM
6616..6653
/note="(SH3BGR)"
/rpt_type=TANDEM
6654..6691
/note="(SH3BGR)"
/rpt_type=TANDEM
6692..6729
/note="(SH3BGR)"
/rpt_type=TANDEM
6730..6767
/note="(SH3BGR)"
/rpt_type=TANDEM
6768..6805
/note="(SH3BGR)"
/rpt_type=TANDEM
6806..6843
/note="(SH3BGR)"
/rpt_type=TANDEM
6844..6881
/note="(SH3BGR)"
/rpt_type=TANDEM
6882..6919
/note="(SH3BGR)"
/rpt_type=TANDEM
6920..6957
/note="(SH3BGR)"
/rpt_type=TANDEM
6958..6995
/note="(SH3BGR)"
/rpt_type=TANDEM
6996..7033
/note="(SH3BGR)"
/rpt_type=TANDEM
7034..7071
/note="(SH3BGR)"
/rpt_type=TANDEM
7072..7109
/note="(SH3BGR)"
/rpt_type=TANDEM
7110..7147
/note="(SH3BGR)"
/rpt_type=TANDEM
7148..7185
/note="(SH3BGR)"
/rpt_type=TANDEM
7186..7223
/note="(SH3BGR)"
/rpt_type=TANDEM
7224..7261
/note="(SH3BGR)"
/rpt_type=TANDEM
7262..7299
/note="(SH3BGR)"
/rpt_type=TANDEM
7300..7337
/note="(SH3BGR)"
/rpt_type=TANDEM
7338..7375
/note="(SH3BGR)"
/rpt_type=TANDEM
7376..7413
/note="(SH3BGR)"
/rpt_type=TANDEM
7414..7451
/note="(SH3BGR)"
/rpt_type=TANDEM
7452..7489
/note="(SH3BGR)"
/rpt_type=TANDEM
7490..7527
/note="(SH3BGR)"
/rpt_type=TANDEM
7528..7565
/note="(SH3BGR)"
/rpt_type=TANDEM
7566..7603
/note="(SH3BGR)"
/rpt_type=TANDEM
7604..7641
/note="(SH3BGR)"
/rpt_type=TANDEM
7642..7679
/note="(SH3BGR)"
/rpt_type=TANDEM
7680..7717
/note="(SH3BGR)"
/rpt_type=TANDEM
7718..7755
/note="(SH3BGR)"
/rpt_type=TANDEM
7756..7793
/note="(SH3BGR)"
/rpt_type=TANDEM
7794..7831
/note="(SH3BGR)"
/rpt_type=TANDEM
7832..7869
/note="(SH3BGR)"
/rpt_type=TANDEM
7870..7907
/note="(SH3BGR)"
/rpt_type=TANDEM
7908..7945
/note="(SH3BGR)"
/rpt_type=TANDEM
7946..7983
/note="(SH3BGR)"
/rpt_type=TANDEM
7984..8021
/note="(SH3BGR)"
/rpt_type=TANDEM
8022..8059
/note="(SH3BGR)"
/rpt_type=TANDEM
8060..8097
/note="(SH3BGR)"
/rpt_type=TANDEM
8098..8135
/note="(SH3BGR)"
/rpt_type=TANDEM
8136..8173
/note="(SH3BGR)"
/rpt_type=TANDEM
8174..8211
/note="(SH3BGR)"
/rpt_type=TANDEM
8212..8249
/note="(SH3BGR)"
/rpt_type=TANDEM
8250..8287
/note="(SH3BGR)"
/rpt_type=TANDEM
8288..8325
/note="(SH3BGR)"
/rpt_type=TANDEM
8326..8363
/note="(SH3BGR)"
/rpt_type=TANDEM
8364..8401
/note="(SH3BGR)"
/rpt_type=TANDEM
8402..8439
/note="(SH3BGR)"
/rpt_type=TANDEM
8440..8477
/note="(SH3BGR)"
/rpt_type=TANDEM
8478..8515
/note="(SH3BGR)"
/rpt_type=TANDEM
8516..8553
/note="(SH3BGR)"
/rpt_type=TANDEM
8554..8591
/note="(SH3BGR)"
/rpt_type=TANDEM
8592..8629
/note="(SH3BGR)"
/rpt_type=TANDEM
8630..8667
/note="(SH3BGR)"
/rpt_type=TANDEM
8668..8705
/note="(SH3BGR)"
/rpt_type=TANDEM
8706..8743
/note="(SH3BGR)"
/rpt_type=TANDEM
8744..8781
/note="(SH3BGR)"
/rpt_type=TANDEM
8782..8819
/note="(SH3BGR)"
/rpt_type=TANDEM
8820..8857
/note="(SH3BGR)"
/rpt_type=TANDEM
8858..8895
/note="(SH3BGR)"
/rpt_type=TANDEM
8896..8933
/note="(SH3BGR)"
/rpt_type=TANDEM
8934..8971
/note="(SH3BGR)"
/rpt_type=TANDEM
8972..9009
/note="(SH3BGR)"
/rpt_type=TANDEM
9010..9047
/note="(SH3BGR)"
/rpt_type=TANDEM
9048..9085
/note="(SH3BGR)"
/rpt_type=TANDEM
9086..9123
/note="(SH3BGR)"
/rpt_type=TANDEM
9124..9161
/note="(SH3BGR)"
/rpt_type=TANDEM
9162..9199
/note="(SH3BGR)"
/rpt_type=TANDEM
9200..9237
/note="(SH3BGR)"
/rpt_type=TANDEM
9238..9275
/note="(SH3BGR)"
/rpt_type=TANDEM
9276..9313
/note="(SH3BGR)"
/rpt_type=TANDEM
9314..9351
/note="(SH3BGR)"
/rpt_type=TANDEM
9352..9389
/note="(SH3BGR)"
/rpt_type=TANDEM
9390..9427
/note="(SH3BGR)"
/rpt_type=TANDEM
9428..9465
/note="(SH3BGR)"
/rpt_type=TANDEM
9466..9503
/note="(SH3BGR)"
/rpt_type=TANDEM
9504..9541
/note="(SH3BGR)"
/rpt_type=TANDEM
9542..9579
/note="(SH3BGR)"
/rpt_type=TANDEM
9580..9617
/note="(SH3BGR)"
/rpt_type=TANDEM
9618..9655
/note="(SH3BGR)"
/rpt_type=TANDEM
9656..9693
/note="(SH3BGR)"
/rpt_type=TANDEM
9694..9731
/note="(SH3BGR)"
/rpt_type=TANDEM
9732..9769
/note="(SH3BGR)"
/rpt_type=TANDEM
9770..9807
/note="(SH3BGR)"
/rpt_type=TANDEM
9808..9845
/note="(SH3BGR)"
/rpt_type=TANDEM
9846..9883
/note="(SH3BGR)"
/rpt_type=TANDEM
9884..9921
/note="(SH3BGR)"
/rpt_type=TANDEM
9922..9959
/note="(SH3BGR)"
/rpt_type=TANDEM
9960..9997
/note="(SH3BGR)"
/rpt_type=TANDEM
```


Mon Apr 21 09:59:17 2003

us-09-719-554-3_copy_5000_5500_1.rge

Page 15

Search completed: April 17, 2003, 06:41:04
Job time : 1888.47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 02:30:00 ; Search time 180.787 Seconds

(without alignments)
6240.778 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501
Sequence: 1 caagatctcagatattacaa.....tcagtgaggaacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2004.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	10499	21	ABN97929
2	501	100.0	56093	24	ABL61744
3	474	94.6	2938	20	AA525656
4	474	94.6	2938	21	AA525656
5	474	94.6	2938	22	AA525656
6	453.8	90.6	3831	23	AA571727
7	442	88.2	7582	20	AA525656
8	442	88.2	7582	21	AA525656
9	435	86.8	3867	23	AA568621

10	435	86.8	3867	23	AA576464	DNA encoding novel
11	435	86.8	3867	23	AA580471	DNA encoding novel
12	435	86.8	4349	23	AA576475	DNA encoding novel
13	435	86.8	4535	23	AA576205	DNA encoding novel
14	435	86.8	5154	23	AA567609	DNA encoding novel
15	428.4	85.5	3372	20	AA525663	Human endogenous r
16	428.4	85.5	3372	21	AA559213	Partial pol gene a
17	422.2	84.3	900	23	AA572228	DNA encoding novel
18	422.2	84.3	900	23	AA572228	DNA encoding novel
19	422.2	84.3	900	23	AA573911	DNA encoding novel
20	422.2	84.3	900	23	AA576470	DNA encoding novel
21	422.2	84.3	1493	23	AA576606	DNA encoding novel
22	422.2	84.3	1687	23	AA572234	DNA encoding novel
23	422.2	84.3	1729	23	AA577018	DNA encoding novel
24	422.2	84.3	1743	23	AA584207	DNA encoding novel
25	422.2	84.3	2304	19	AA543199	Multiple sclerosis
26	420.6	84.0	924	23	AA572224	DNA encoding novel
27	420.6	84.0	1660	23	AA573608	DNA encoding novel
28	420.6	84.0	1660	23	AA576459	DNA encoding novel
29	420.6	84.0	1660	23	AA584201	DNA encoding novel
30	420.6	84.0	1810	23	AA585801	DNA encoding novel
31	420.6	84.0	1810	23	AA586023	DNA encoding novel
32	420.6	84.0	1846	23	AA573605	DNA encoding novel
33	420.6	84.0	1846	23	AA584192	DNA encoding novel
34	420.6	84.0	2220	23	AA567604	DNA encoding novel
35	420.6	84.0	2220	23	AA573919	DNA encoding novel
36	420.6	84.0	2220	23	AA576473	DNA encoding novel
37	419.6	83.8	1158	16	AA572815	Multiple sclerosis
38	419.6	83.8	1158	17	AA579793	Partial sequence o
39	419.6	83.8	1158	18	AA576527	MSRV-1B pol.* seque
40	419.6	83.8	1158	19	AA543115	Multiple sclerosis
41	419.6	83.8	2365	19	AA543200	Multiple sclerosis
42	419.6	83.8	2391	19	AA576478	MSRV-1 pol gene se
43	419.6	83.8	2391	19	AA543166	Multiple sclerosis
44	419	83.6	1117	23	AA576460	DNA encoding novel
45	419	83.6	1389	23	AA567594	DNA encoding novel

ALIGNMENTS

RESULT 1
ID ABN97929 standard; DNA; 10499 BP.
AC ABN97929;
DT 01-AUG-2002 (first entry)
DE Human retroviral sequence HERV-7g.
KW Autoimmune disease; HERV-7g; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
OS Human retrovirus.
PN WO9967395-A1.
PD 29-DEC-1999.
PF 23-JUN-1999; 99WO-FR01513.
PR 23-JUN-1998; 98FR-0007920.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI Alliel PM, Perin J, Rieger F;
DR WPI; 2000-160587/14.
FT New nucleic acid sequences of human endogenous retrovirus, HERV-7g,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -

XX Claim 3; Fig 1; 225bp; French.
PS
XX
CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;

Query Match 100.0%; Score 501; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 1.6e-151;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGGATTATCAATGAGCTGTTGTTCTCTATAGCCAGCTTACCTGACCT 60
DB 5000 CAAGATCTCAGGATTATCAATGAGCTGTTGTTCTCTATAGCCAGCTTACCTGACCT 5059
QY 61 TATACCTGCTTCCCAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 5060 TATACCTGCTTCCCAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
QY 121 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCTG 180
DB 5120 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 5179
QY 181 ACTTGAACCCCAATCTCACTCACTGAGCTTTTATCCCAAGGCTTCAAGGATAGT 240
DB 5180 ACTTGAACCCCAATCTCACTCACTGAGCTTTTATCCCAAGGCTTCAAGGATAGT 5239
QY 241 CCCCACTATTTGGCCAGGATTAAGCCCAAGACTGAGCAATCCCAATCCCAATCCCAATCC 300
DB 5240 CCCCACTATTTGGCCAGGATTAAGCCCAAGACTGAGCAATCCCAATCCCAATCCCAATCC 5299
QY 301 TGTCTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 5300 TGTCTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5359
QY 361 AGCCACCCCAAGGCTTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 5360 AGCCACCCCAAGGCTTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5419
QY 421 GGCTCAACTCTGCTCAACAGAGGTTACTTAAGGCTTAATTTCCAAAGCCAGGAGGCC 480
DB 5420 GGCTCAACTCTGCTCAACAGAGGTTACTTAAGGCTTAATTTCCAAAGCCAGGAGGCC 5479
QY 481 CTCAGTGAAGAACATCCAG 501
DB 5480 CTCAGTGAAGAACATCCAG 5500

KW gene; ds.
XX
OS Homo sapiens.
XX
PM WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-235133P.
PR 18-SEP-2000; 2000US-235617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.

(AVAIL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Sopet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.

Screening for anti-neoplastic agent involves exposing cells to a
PR chemical agent to be tested for anti-neoplastic activity, and
PR determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 81; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL6164
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic

AC AAS9206;
XX 07-NOV-2000 (first entry)
XX
DE Gag and partial pol gene fragment of HERV-W from human genome.
XX
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
XX gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS Homo sapiens.
XX MO200043521-A2.
XX
XX 27-JUL-2000.
XX
XX 21-JAN-2000; 2000MO-FR00144.
XX PF
XX 21-JAN-1999; 99FR-0000889.
XX
XX (INMR) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Mallet F, Voisset C;
XX WPI; 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
XX diagnosis of autoimmune disease and complications of pregnancy.
XX PT contains at least part of the gag gene -
XX
XX Disclosure; Page 43; 53pp; French.
XX
XX The present sequence represents an endogenous retroviral nucleic acid
XX fragment, which is associated with an autoimmune disease, and is
XX integrated into the human genome. The fragment is originally derived
XX from a novel retrovirus, human endogenous retrovirus W (HERV-W). The
XX HERV-W retrovirus is associated with autoimmune disease, failure of
XX pregnancy or disorders of pregnancy. The nucleic acid fragment, or
XX proteins derived from it, are useful for diagnosis of autoimmune
XX disease (specifically multiple sclerosis) and for monitoring pregnancy.
XX The nucleic acid fragments may also be used for in situ labelling of
XX isolated chromosomes, while the transcription product can be used to
XX study or monitor T cell proliferation in vitro.
XX
SQ Sequence 2938 BP; 878 A; 720 C; 646 G; 692 T; 2 other;

Query Match 94.6%; Score 474; DB 21; Length 2938;
Best Local Similarity 97.8%; Pred. No. 5e-143;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 CAGATCTCAGGATTATCAAT-GAGGCTGTGTCTCTATAGCAGCTGTAAGCC 59
DB 2079 CAGATCTCAGGATTATCAATGAGGCTGTGTCTCTATAGCAGCTGTAAGCC 2138
QY 60 TTATACCTGCTTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
DB 2139 TTATACCTGCTTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
QY 120 GGATGCTTCTTCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCG 179
DB 2199 GGATGCTTCTTCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCG 2258
QY 180 TACTTCAACCCAGATCTCACTCACTGACATTTTACCCCAAGGGTTAGGGATAG 239
DB 2259 TACTTCAACCCAGATCTCACTCACTGACATTTTACCCCAAGGGTTAGGGATAG 2318
QY 240 TCCCATCTATTGGCCAGGATTAAGCCCAAGATTAGCCCAATCTCTATCTGAGAC 299
DB 2319 CCCCATCTATTGGCCAGGATTAAGCCCAAGATTAGCCCAATCTCTATCTGAGAC 2378
QY 300 TTGTCCTTGGAGAGTGAATTAATTTTGGCGCCCATTCAGAAACCTTGCCATC 359
DB 2379 TTGTCCTTGGAGAGTGAATTAATTTTGGCGCCCATTCAGAAACCTTGCCATC 2438

QY 360 AAGCCACCAAGCGCTTCAATTTCTGCTACCTGTGCTACATGATTTCCAAACCA 419
DB 2439 AAGCCACCAAGCGCTTCAATTTCTGCTACCTGTGCTACATGATTTCCAAACCA 2498
QY 420 AGGCTCACTCTGCTCAGAGAGTTACTTAGGGCTAAATATCCAAAGCACCAGGCG 479
DB 2499 AGGCTCACTCTGCTCAGAGAGTTACTTAGGGCTAAATATCCAAAGCACCAGGCG 2558
QY 480 CCTCAGTGGAGACACATCCAG 501
DB 2559 CCTCAGTGGAGACACATCCAG 2580

RESULT 5
ID AAS68626 standard; cDNA; 7466 BP.
AC AAS68626;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #4430.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSSE-) HYSSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG04439.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID No 4430; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy involving
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensic, gene mapping, identification of mutations in
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.

SQ Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;
Query Match 94.6%; Score 474; DB 23; Length 7466;
Best Local Similarity 97.8%; Pred. No. 7,66-143;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 1 CAAGATCTCAGAGTATTAAT- GAGGCTGTGTTCCTTATAGCCAGCTGTACCTAGCCC 59
DB 6607 CAAATCTCAGAGTATTAATCAATGAGAGCGCTGTCTTATATACCAGCTGTACCTAGCCC 6666
QY 60 TTATATCTGTCTTCCCAATATCCAGAGAGAGAGAGGTTTACAGTCCCTGACCTTCA 119
DB 6667 TTATATCTGTCTTCCCAATATCCAGAGAGAGAGAGGTTTACAGTCCCTGACCTTCA 6726
QY 120 GGAATGCTTCTTCTGATCCCTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTG 179
DB 6727 GGAATGCTTCTTCTGATCCCTGTGATCTGTGATCTGTGATCTGTGATCTGTGATCTGTG 6786
QY 180 TACTTCAACCCCAATCTCACTCACTGACCTGATCTATTTTACCCCAAGGTTCAAGGATAG 239
DB 6787 TACTTCAACCCCAATCTCACTCACTGACCTGATCTATTTTACCCCAAGGTTCAAGGATAG 6846
QY 240 TCCCATCTATTTTGGCCAGGCTTATAGCCCAAGCTTGAAGCTTCTATCTGTGATC 299
DB 6847 CCCCCTATTTTGGCCAGGCTTATAGCCCAAGCTTGAAGCTTCTATCTGTGATC 6906
QY 300 TTGCTCTTGGAGTGTGATGATTTACTTTTGGCCGCTTCAAGAACTTGTGATC 359
DB 6907 TTGCTCTTGGAGTGTGATGATTTACTTTTGGCCGCTTCAAGAACTTGTGATC 6966
QY 360 AAGCACCACCAAGCGCTCTTCAATTTCTGCTACCTGTGAGCTTCAAGGTTTCAAAACCA 419
DB 6967 AAGCACCACCAAGCGCTCTTCAATTTCTGCTACCTGTGAGCTTCAAGGTTTCAAAACCA 7026
QY 420 AGGCTCAACTCTGCTCAACAGAGTACTTATAGGCTTAAATATTCAAAGGCACTGAGGC 479
DB 7027 AGGCTCAACTCTGCTCAACAGAGTACTTATAGGCTTAAATATTCAAAGGCACTGAGGC 7086
QY 480 CCTCAGTAGGAACATCTCCAG 501
DB 7087 CCTCAGTAGGAACATCTCCAG 7108
RESULT 6
ID AAS71727 standard; cDNA; 3831 BP.
AC AAS71727;
DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #7531.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX PD 30-MAR-2001; 2001WO-US08631.
XX PF 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG07540.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 7531; 103bp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
XX polynucleotide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pcr_sequences](http://wipo.int/pub/published_pcr_sequences).
XX
SQ Sequence 3831 BP; 1173 A; 953 C; 815 G; 890 T; 0 other;
Query Match 90.6%; Score 453.8; DB 23; Length 3831;
Best Local Similarity 95.2%; Pred. No. 26-136;
Matches 479; Conservative 0; Mismatches 22; Indels 2; Gaps 1;
QY 1 CAAGATCTCAGAGTATTAATCAATGAGGCTGTGCTTCTATATAGCCAGCTGTACCTAGCCCT 60
DB 412 CAAATCTCAGAGTATTAATCAATGAGGCTGTGCTTCTATATAGCCAGCTGTACCTAGCCCT 471
QY 61 TATATCTGCTTCCCAATATCCAGAGAGAGAGGTTTATAGTCTGTGATC 120
DB 472 TATATCTGCTTCCCAATATCCAGAGAGAGAGGTTTATAGTCTGTGATC 531
QY 121 GATGCTTCTTCTGATCCCGTACCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 180
DB 532 GATGCTTCTTCTGATCCCGTACCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 591
QY 181 ACTTCAACCCCAATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 240
DB 592 CTTTCAACCCCAATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 651
QY 241 CCCCCTATTTTGGCCAGGCTTATAGCCCAAGCTTGAAGCTTCTATCTGTGATC 298
DB 652 CTCATCTATTTTGGCCAGGCTTATAGCCCAAGCTTGAAGCTTCTATCTGTGATC 711
QY 299 CTTGCTCTGCTGATGATGATTTACTTTTGGCCGCTTCAAGAACTTGTGATC 358
DB 712 CTTGCTCTGCTGATGATGATTTACTTTTGAAGCACTTGAAGAACTTGTGATC 771
QY 359 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGATGTTTCCAAACA 418
DB 772 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGATGTTTCCAAACA 831
QY 419 AAGGCTCACTCTGCTCAACAGAGTACTTATAGGCTTAAATATTCAAAGGCACTGAGGC 478
DB 832 AAGGCTCACTCTGCTCAACAGAGTACTTATAGGCTTAAATATTCAAAGGCACTGAGGC 891
QY 479 CCTCAGTAGGAACATCTCCAG 501
DB 892 CCTCAGTAGGAACATCTCCAG 914

```

RESULT 7
AA25665
ID   AAX25665 standard; cDNA to mRNA; 7582 BP.
XX
XX   AAX25665;
XX
XX   21-MAY-1999 (first entry)
XX
XX   Complete human endogenous retrovirus W genome.
XX
XX   Clone; human endogenous retrovirus; genome; autoimmune disease;
XX   multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
XX   disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX
XX   Human endogenous retrovirus.
XX
XX   MO9902696-A1.
XX
XX   21-JAN-1999.
XX
XX   06-JUL-1998; 98WO-FR01442.
XX
XX   07-JUL-1997; 97FR-0008815.
XX
XX   (IMMR ) BIO MERIEUX.
XX
XX   Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;
XX
XX   WPI; 1999-120897/10.
XX
XX   New nucleic acid sequences from human endogenous retrovirus-W -
XX   expressed exclusively in placenta and useful in diagnosis and
XX   therapy of autoimmune disease, and abnormal or failed pregnancy
XX
XX   Claim 1; Page 71-74; 106pp; French.
XX
XX   This sequence represents the complete sequence of the human endogenous
XX   retrovirus (HERV) W genome. The nucleic acids, their fragments or
XX   peptides encoded by them are markers of autoimmune disease (e.g. multiple
XX   sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
XX   insulin-dependent diabetes and related pathologies) and of abnormal or
XX   unsuccessful pregnancy and can be used as chromosomal markers for
XX   susceptibility to these conditions, or proximity markers of genes
XX   associated with this susceptibility.
XX
XX   Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;
XX
XX   Query Match      88.2%; Score 442; DB 20; Length 7582;
XX   Best Local Similarity 91.3%; Pred. No. 1.8e-132;
XX   Matches 464; Conservative 31; Mismatch:hes 6; Indels 7; Gaps 3;
XX
OY   1 CAAGATTCAGGATTATCAAT-GAGGCTGTGTCCTGTATAGCCAGCTGACTAGCCC 59
DB   2678 CAAGATTCAGGATTATCAATGAGGCCCTGTGTCCTTTATACCAAGCTGACTAGCCC 2737
OY   60 TTATACCTGCTTCCCAATACAGAGAGACAGAGTGTTTACAGTCTGACCTTCA 119
DB   2738 TTATACCTGCTTCCCAATACAGAGAGACAGAGTGTTTACAGTCTGACCTTCA 2797
OY   120 GGATGCTTCTTGTGATCCCTGTACATCTGACTTCAATTTCTTGTGCTTTGAAGA 179
DB   2738 TTATACCTGCTTCCCAATACAGAGAGACAGAGTGTTTACAGTCTGACCTTCA 2797
OY   2738 GGATGCTTCTTGTGATCCCTGTACATCTGACTTCAATTTCTTGTGCTTTGAAGA 2857
DB   2738 GGATGCTTCTTGTGATCCCTGTACATCTGACTTCAATTTCTTGTGCTTTGAAGA 2857
OY   180 TACTTCAAAACCAATCACTCACTGACGATTTTACCCCAAGGTTCAAGGATAG 239
DB   2858 TACTTCAAAACCAATCACTCACTGACGATTTTACCCCAAGGTTCAAGGATAG 2917
OY   240 TCCCATCTATTGGCCAGCATTAAGCCCAAGATTGACCAATCTCATACCTGACA- 298
DB   2918 TCCCATCTATTGGCCAGCATTAAGCCCAAGATTGACCAATCTCATACCTGACA- 2977
OY   299 -CTTGCTCTGGTAGGTGATGATTACTTTTGGCCGCCCATTCAGAACTTGTGCA 357
DB   299 -CTTGCTCTGGTAGGTGATGATTACTTTTGGCCGCCCATTCAGAACTTGTGCA 357

```

```

DB   2978 TCTGTCTCTCRGAKGTGATGATTACTTTTTCGTCGCCYCCYRTTCAGAAACCTGTGCCA 3037
OY   358 TCAAGCCACCAACGCTCTTCAATTTCTCTGCTACTGTGGCTACATGGTTTCCAAACC 417
DB   3038 TCAAGCCACCAACGCTCTTCAATTTCTCTGCTACTGTGGCTACATGGTTTCCAAACS 3097
OY   418 AAAGCTCACTCTGCTCACAGCAGT----TACTTAGGGCTAAATATCCAAAGGCAC 473
DB   3098 ARAAGCTCACTCTGCTCACAGCAGTAAATCTTAGGGCTAAATATTCAAAGGCAC 3157
OY   474 CAGGCGCTCAGTGAAGACATCCAG 501
DB   3158 CAGGCGCTCAGTGAAGACATCCAG 3185

RESULT 8
AA59215
ID   AA59215 standard; DNA; 7582 BP.
XX
XX   AA59215;
XX
XX   07-NOV-2000 (first entry)
XX
XX   Human endogenous retrovirus W (HERV-W) sequence.
XX
XX   Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
XX   gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
XX   Human endogenous retrovirus.
XX
XX   Location/Qualifiers
XX   FH   Key 1..120
XX   FT   LTR
XX   FT   /tag= a
XX   FT   /note= "R of 5' LTR"
XX   FT   LTR 121..575
XX   FT   /tag= b
XX   FT   /note= "US of 5' LTR"
XX   FT   primer_bind 579..596
XX   FT   /tag= c
XX   FT   CDS 5581..7194
XX   FT   /tag= d
XX   FT   /note= "ORF1 env538"
XX   FT   CDS 7039..7194
XX   FT   /tag= e
XX   FT   /note= "ORF2 52 AA"
XX   FT   CDS 7112..7255
XX   FT   /tag= f
XX   FT   /note= "ORF3 48 AA"
XX   FT   misc_feature 7244..7254
XX   FT   /tag= g
XX   FT   /note= "polypurine tract"
XX   FT   LTR 7256..7582
XX   FT   /tag= h
XX   FT   /note= "U3-R of 3' LTR"
XX   FT   polyA_signal 7563..7569
XX   FT   /tag= i
XX   PN   MO200043521-A2.
XX   PD   27-JUL-2000.
XX   PE   21-JAN-2000; 2000WO-FR00144.
XX   PR   21-JAN-1999; 99FR-0000888.
XX   PA   (IMMR ) BIO MERIEUX.
XX   PI   Paranhos-Baccala G, Mallet F, Voisset C;
XX   DR   WPI; 2000-499229/44.
XX
XX   New nucleic acid from human endogenous retrovirus, useful e.g. for
XX   diagnosis of autoimmune disease and complications of pregnancy.

```

PT contains at least part of the gag gene -
XX
PS Disclosure; Page 49-52; 53pp; French.
XX
CC The present sequence represents an endogenous retrovirus, which is
CC associated with an autoimmune disease, and is integrated into the human
CC genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
CC HERV-W retrovirus is associated with autoimmune disease, failure of
CC pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
CC proteins derived from it, are useful for diagnosis of autoimmune
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.
CC The nucleic acid fragments may also be used for in situ labelling of
CC isolated chromosomes, while the transcription product can be used to
CC study or monitor T cell proliferation in vitro.
XX
SQ Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 216 other;

Query Match 88.2%; Score 442; DB 21; Length 7582;
Best Local Similarity 91.3%; Pred. No. 1.8e-132;
Matches 464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;

QY 1 CAAGTCTCAGGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGTGAACCTGAGCC 59
DB 2678 CAAGTCTCAGGATTATCAATGAGGCGGTGTCTTTATACCAGCTGATCCTAGCC 2737
QY 60 TTATCTCTGCTTCCCAATACAGAGAGAGAGTGTTCAGTCTGACCTCA 119
DB 2738 TTATCTCTGTTTCCCAATACAGAGAGAGAGTGTTCAGTCTGACCTCA 2797
QY 120 GAGTCTCTTCTTCTGATCCCTGATACCTGATCTCAATCTGTTGCTTGAAGA 179
DB 2798 GAGTCTCTTCTTCTGATCCCTGATACCTGATCTCAATCTGTTGCTTGAAGA 2857
QY 180 TACTTCAACCAACATCTCAACTCACTGACATTTTACCCCAAGGTTCAAGGAAG 239
DB 2858 TACTTCAACCAACATCTCAACTCACTGACATTTTACCCCAAGGTTCAAGGAAG 2917
QY 240 TCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGACCAATCTGATCTGAGCA- 298
DB 2918 TCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGACCAATCTGATCTGAGCA- 2977
QY 299 -CTTGTCTTGGTAGTGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGCCA 357
DB 2978 TCTGTCTTGGTAGTGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGCCA 3037
QY 358 TCAAGCCACCCAGGCTCTTCAATTTCTGCTACCTGTGCTACAGTTTCCAAAC 417
DB 3038 TCAAGCCACCCAGGCTCTTCAATTTCTGCTACCTGTGCTACAGTTTCCAAAC 3097
QY 418 AAAGCTCAACTGTGCTACAGCAGAGT---TACTTAGGCTAAATATCCAAAGGAC 473
DB 3098 AAAGCTCAACTGTGCTACAGCAGAGT---TACTTAGGCTAAATATCCAAAGGAC 3157
QY 474 CAGGCGCTCAGTGAGAACACATCCAG 501
DB 3158 CAGGCGCTCAGTGAGAACACATCCAG 3185

RESULT 9
AAS68621
ID AAS68621 standard; cDNA; 3867 BP.
XX
XX AAS68621;
XX
XX 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4425.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX

PN WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
DR P-PSDB; ABG04434.
XX
XX WPI; 2001-639362/73.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 4425; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantifying a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3867 BP; 1148 A; 992 C; 838 G; 869 T; 0 other;

Query Match 86.8%; Score 435; DB 23; Length 3867;
Best Local Similarity 93.9%; Pred. No. 2.5e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

QY 1 CAATATCTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGTGAACCTGAGCC 60
DB 1693 CAATATCTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGTGAACCTGAGCC 1752
QY 61 TATATCTGCTTCCCAATACAGAGAGAGAGTGTTCCTATACCAAGTATACCTGAGCC 120
DB 1753 TATATCTGCTTCCCAATACAGAGAGAGAGTGTTCCTATACCAAGTATACCTGAGCC 1812
QY 121 GATGCTTCTTCTGATCCCTGATACCTGATCTCAATCTGTTGCTTGAAGAT 180
DB 1813 GATGCTTCTTCTGATCCCTGATACCTGATCTCAATCTGTTGCTTGAAGAT 1872
QY 181 ACTTCAAAACCAATCTCAACTGACCTGACATTTTACCCCAAGGTTCAAGGATAGT 240
DB 1873 ACTTCAAAACCAATCTCAACTGACCTGACATTTTACCCCAAGGTTCAAGGATAGT 1932
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAGCCCAATCCTGACACT 300
DB 1933 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAGCCCAATCCTGACACT 1992
QY 301 --TGTCTTGGTAGTGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGCTAT 358
|||||

Db 1993 CCTGCTCCTCAGTGCATGATGATTACTTTTA3CTGCCCGTTCAAGAACTTGGCCAT 2052
Qy 359 CAAGCCACCCAAAGCGCTCTTCAATTTCTCGTACCTGTGCTACATGCTTTCCAAACCA 418
Db 2053 CAAGCCACCCAAAGCGCTCTTCAATTTCTCGTACCTGTGCTACATGCTTTCCAAACCA 2112
Qy 419 AAGGCTCAACTCTGCTCAACAGAGTTA---CTTAGGGCTTAAATTTTCCAAAGGCACC 474
Db 2113 AAGGCTCAACTCTGCTCAACAGAGTTAATGCTTAGGGCTTAAATTTTCCAAAGTCACC 2172
Qy 475 AGGGCCCTCAGTAGAGAAACATCCAG 501
Db 2173 AGGGCCCTCAGTAGAGAAACGTAATCCAG 2199

RESULT 10
AAS76464
ID AAS76464 standard; cDNA, 3867 BP.
XX
AC AAS76464;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #12268.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG12277.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 1; SEQ ID No 12268; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, its molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ffp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;
Query Match 86.8%; Score 435; DB 23; Length 3867;
Best local Similarity 93.9%; Pred. No. 2,5e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
Qy 1 CAAGATCTCAGGATTAATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACTAGCCCT 60
Db 1693 CAAGATCTCAGGATTAATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACTAGCCCT 1752
Qy 61 TATCTCTGCTTTCCCAAATACAGAGAGAGAGAGGTTTACAGTCCGAGACCTTCAG 120
Db 1753 TATCTCTGCTTTCCCAAATACAGAGAGAGAGAGGTTTACAGTCCGAGACCTTCAG 1812
Qy 121 GATGCTTCTTCTGTCATCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAGAT 180
Db 1813 GATGCTTCTTCTGTCATCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAGAT 1872
Qy 181 ACTTCAAAACCAACATCTCAATCTGACCTGACCTATTTTACCCCAAGGCTTCAGGATAGT 240
Db 1873 CCTTCAAAACCAACATCTCAATCTGACCTGACCTGATTTTACCCCAAGGCTTCAGGATAGT 1932
Qy 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAGCCAAATCTCATACCTGGACACT 300
Db 1933 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAGCCAAATCTCATACCTGGACACT 1992
Qy 301 --TGTCCTGGGTAGTGATGATTTACTTTTGGCCGCCCATTCAGAAACCTTGTGCAT 358
Db 1993 CCTGCTCTTCAAGTCATGATGATTTACTTTTGGCCGCCCATTCAGAAACCTTGTGCAT 2052
Qy 359 CAAGCCACCCAAAGCGCTCTTCAATTTCTCGTACCTGTGCTACATGCTTTCCAAACCA 418
Db 2053 CAAGCCACCCAAAGCGCTCTTCAATTTCTCGTACCTGTGCTACATGCTTTCCAAACCA 2112
Qy 419 AAGGCTCAACTCTGCTCAACAGAGTTA---CTTAGGGCTTAAATTTTCCAAAGGCACC 474
Db 2113 AAGGCTCAACTCTGCTCAACAGAGTTAATGCTTAGGGCTTAAATTTTCCAAAGTCACC 2172
Qy 475 AGGGCCCTCAGTAGAGAAACATCCAG 501
Db 2173 AGGGCCCTCAGTAGAGAAACGTAATCCAG 2199

RESULT 11
AAS80471
ID AAS80471 standard; cDNA, 3867 BP.
XX
AC AAS80471;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #16275.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI, 2001-639362/73.
DR P-PSDB; ABG16284.
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
XX Claim 1; SEQ ID No 16275; 103bp; English.
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;
Query Match 86.8%; Score 435; DB 23; Length 3867;
Best Local Similarity 93.9%; Pred. No. 2.5e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
QY 1 CAAGATCTCAGGATTAATCAATGAGGCTGTGTCCTCATATGCGCAGCTGATACGCTT 60
DB 1693 CAAGATCTCAGGATTAATCAATGAGGCTGTGTCCTCATATGCGCAGCTGATACGCTT 1752
QY 61 TATACCTGCTTTCCCAATATCAGAGAGACAGAGTGTATACAGTCTGAGACCTTCAG 120
DB 1753 TATACCTGCTTTCCCAATATCAGAGAGACAGAGTGTATACAGTCTGAGACCTTCAG 1812
QY 121 GATGCTCTTCTTGATCCTCTGATCCTGATCTCAATCTTCTGTTGCTTGAAGAT 180
DB 1813 GATGCTCTTCTTGATCCTCTGATCCTGATCTCAATCTTCTGTTGCTTGAAGAT 1872
QY 181 ACTTCAACCCCAATCTCACTCAGCCTGAGCTATTTTACCACCAAGGTTGAGGATAGT 240
DB 1873 CTTTCAACCCCAATCTCACTCAGCCTGAGCTATTTTACCACCAAGGTTGAGGATAGT 1932
QY 241 CCCCATCTATTGGCCAGGATTAAGCCCAAGACTTGAGCCATCTTACCTGAGCACT 300
DB 1933 CCCCATCTATTGGCCAGGATTAAGCCCAAGACTTGAGCCATCTTACCTGAGCACT 1992
QY 301 --TGTCTTGGTATGATGATTTACTTTTGGCCGCCCATTTCAAAAACCTTGTGCCAT 358
DB 1993 CTTGTCTTGGTATGATGATTTACTTTTGGCCGCCCATTTCAAAAACCTTGTGCCAT 2052
QY 359 CAAGCCACCCCAAGGCTTTCATTTCTCGCTACCTGTGCTACATGCTTTCCAAACCA 418
DB 2053 CAAGCCACCCCAAGGCTTTCATTTCTCGCTACCTGTGCTACATGCTTTCCAAACCA 2112
QY 419 AAGGCTCAACTCTGCTCAACAGGTTA---CTTAGGGCTAAATTTCCAAAGGCACC 474
DB 2113 AAGGCTCAACTCTGCTCAACAGGTTA---CTTAGGGCTAAATTTCCAAAGGCACC 2172
QY 475 AGGGCCCTCACTGAGGAACACATCCAG 501
DB 2173 AGGGCCCTCACTGAGGAACATCCAG 2199

RESULT 12
AAS76475
ID AAS76475 standard; cDNA; 4349 BP.
XX
XX AAS76475;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
DE DNA encoding novel human diagnostic protein #12279.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
XX W0200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA
PI Dmanac RT, Liu C, Tang YT;
PT
XX
XX WPI, 2001-639362/73.
DR P-PSDB; ABG12288.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
XX Claim 1; SEQ ID No 12279; 103bp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4349 BP; 1319 A; 1061 C; 941 G; 1026 T; 2 other;
Query Match 86.8%; Score 435; DB 23; Length 4349;
Best Local Similarity 93.9%; Pred. No. 2.6e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
QY 1 CAAGATCTCAGGATTAATCAATGAGGCTGTGTCCTCATATGCGCAGCTGATACGCTT 60
DB 2382 CAAGATCTCAGGATTAATCAATGAGGCTGTGTCCTCATATGCGCAGCTGATACGCTT 2441
QY 61 TATACCTGCTTTCCCAATATCAGAGAGACAGAGTGTATACAGTCTGAGACCTTCAG 120

```
Db 2442 TATACCTGCTTCCCAATACCAAGAGAACAAAGTGTTTACAGTCTGATCTTAAG 2501
Qy 121 GATGCTCTCTGATGATCCGTGATCATCTGATCTCAATCTGTTTGGCTTGAAGAT 180
Db 2502 GATGCTCTCTGATGATCCGTGATCATCTGATCTCAATCTGTTTGGCTTGAAGAT 2561
Qy 181 ACTTCAACCAACATCTCAATCACTGATCTATTTTACCACAGGTTGAGGATAGT 240
Db 2562 CCTTCAACCAACATCTCAATCACTGATCTATTTTACCACAGGTTGAGGATAGT 2621
Qy 241 CCCCATCTATTTGGCCAGGACATTAAGCCCAAGCTTGAACCAATCTCTATCTGACACT 300
Db 2622 CCCCATCTATTTGGCCAGGACATTAAGCCCAAGCTTGAACCAATCTCTATCTGACACT 2681
Qy 301 --TGTCCTCGGTAGTGGATGATTACTTTTGACCGCCCATTTGAGAACCTTGCCAT 358
Db 2682 CCGTCTCTTCAAGTGAATGATGATTTACTTTTACCTGCCCCCTTCAGAACCTTGTCAT 2741
Qy 359 CAAGCCACCCAAAGCGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 418
Db 2742 CAAGCCACCCAAAGCGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 2801
Qy 419 AAGGCTCACTCTGCTCAACAGAGTTA---CTTAGGGCTTAAATTTCCAAAGGCACC 474
Db 2802 AAGGCTCACTCTGCTCAACAGAGTTAATGCTTAGGGCTTAAATTTCCAAAGTCACC 2861
Qy 475 AGGGCCCTCAGTAGAAGAACATCCAG 501
Db 2862 AGGGCCCTCAGTAGAAGAACATCCAG 2888

RESULT 13
AAS76205
ID AAS76205 standard; cDNA, 4535 BP.
XX
AC AAS76205;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #12009.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PsDB; ABG12018.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX
XX Claim 1; SEQ ID No 12009; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
```

```
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4535 BP; 1356 A; 1159 C; 983 G; 1037 T; 0 other;
XX
Query Match 86.8%; Score 435; DB 23; Length 4535;
Best Local Similarity 93.9%; Pred. No. 2, 7e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

Qy 1 CAAGATTCAGGATTAATCAATGAGGCTGTTGTTCTTATAGCAAGTGTACTAGCCCT 60
Db 1693 CAAGATTCAGGATTAATCAATGAGGCTGTTGTTCTTATAGCAAGTGTACTAGCCCT 1752
Qy 61 TATACCTGCTTCCCAATACCAAGAGAGAGGTTTACAGTCTGACCTTCAAG 120
Db 1753 TATACCTGCTTCCCAATACCAAGAGAGAGGTTTACAGTCTGACCTTCAAG 1812
Qy 121 GATGCTCTCTGATGATCCGTGATCATCTGATCTCAATCTGTTTGGCTTGAAGAT 180
Db 1813 GATGCTCTCTGATGATCCGTGATCATCTGATCTCAATCTGTTTGGCTTGAAGAT 1872
Qy 181 ACTTCAACCAACATCTCAATCACTGATCTATTTTACCACAGGTTGAGGATAGT 240
Db 1873 ACTTCAACCAACATCTCAATCACTGATCTATTTTACCACAGGTTGAGGATAGT 1932
Qy 241 CCCCATCTATTTGGCCAGGACATTAAGCCCAAGCTTGAACCAATCTCTATCTGACACT 300
Db 1933 CCCCATCTATTTGGCCAGGACATTAAGCCCAAGCTTGAACCAATCTCTATCTGACACT 1992
Qy 301 --TGTCCTCGGTAGTGGATGATTACTTTTGACCGCCCATTTGAGAACCTTGTCAT 358
Db 1993 CCGTCTCTTCAAGTGAATGATTTACTTTTACTGCTCCGTTGAGAACCTTGTCAT 2052
Qy 359 CAAGCCACCCAAAGCGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 418
Db 2053 CAAGCCACCCAAAGCGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 2112
Qy 419 AAGGCTCACTCTGCTCAACAGAGTTA---CTTAGGGCTTAAATTTCCAAAGGCACC 474
Db 2113 AAGGCTCACTCTGCTCAACAGAGTTAATGCTTAGGGCTTAAATTTCCAAAGTCACC 2172
Qy 475 AGGGCCCTCAGTAGAAGAACATCCAG 501
Db 2173 AGGGCCCTCAGTAGAAGAACATCCAG 2199

RESULT 14
AAS67609
ID AAS67609 standard; cDNA, 5154 BP.
XX
AC AAS67609;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #3413.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
```

XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
XX	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSEQ-) HYSEQ INC.
XX	
PI	Drimanac RT, Liu C, Tang YT;
XX	
DR	WPI: 2001-639362/73.
XX	
DR	P-PSDB: ABG03422.
XX	
PT	New isolated polypeptide and encoded polypeptides, useful in
XX	diagnostics, forensics, gene mapping, identification of mutations
XX	responsible for genetic disorders or other traits and to assess
PT	biodiversity
XX	
PS	Claim 1; SEQ ID No 3413; 103bp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences, (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIFO
CC	at ftp.wifo.int/pub/published_pct_sequences.
XX	
XX	Sequence 5154 BP; 1537 A; 1321 C; 1117 G; 1179 T; 0 other;
XX	
Query Match	86.8%; Score 435; DB 23; Length 5154;
Best Local Similarity	93.9%; Pred. No. 2.9e-130;
Matches 476; Conservative	0; Mismatches 25; Indels 6; Gaps 2.
1	CAAGATCTCAGATTATCATGAGCGTGTTCCTCTATACCCAGCTGTACCTGACCCCT 60
1693	CAAGATCTCAGATTATCATGAGCGTGTTCCTCTATACCCAGCTGTACCTGACCCCT 1752
61	TATACTCTGCTTCCCAAAATACCAAGAGAGAGAGTGTATTACAGTCTCGACCTTGG 120
1753	TATACTCTGCTTCCCAAAATACCAAGAGAGAGAGTGTATTACAGTCTCGACCTTGG 1812
121	GATGCGCTTCTTGATGATCCCTGTATACCTGTGACTCTCAATTTCTGTGGCTTTGAAGT 180
1813	GATGCGCTTCTTGATGATCCCTGTATACCTGTGACTCTCAATTTCTGTGGCAATTTGAAGT 1872
181	ACTTCAAAACCAACATCTCAACTCACTGACTGATTTTAAACCCCAAGGGTTCAGGATAGT 240
1873	CCTTCAAAACCAACATCTCAACTCACTGACTGATTTTAAACCCCAAGGGTTCAGGATAGT 1932
241	CCCACATATTATTTGGCCAGGATTTAGCCCAAGACTTGAGCCAAATCTCATATCCTGGAACCT 300

Db	1933	CCCCATCTATTTCGGCCAGGACATTTAGCCCAAGACTTGAGCCAGTTCTCATCTCGACACT	1992		
Qy	301	-TGTCTTGGGAGTGGATGATTTACTTTTGGCCGCCCATTCAGAAACCTTGTCAT	358		
Db	1993	CTGTCTTCTAGTGCATGATGATTAATTACTTTAGTGCGCCGTTGAGAAACCTTGTCAT	2052		
Qy	359	CAGGACACCAAGGCTCTTTCAATTTCTCGTACTCTGTGGTATATGGTTTCCAAACA	418		
Db	2053	CAAGCACCACCAAGCCTTTAAATTTCTCTCACTTACCTGTGGCTACAAAGTTTCCAAACA	2112		
Qy	419	AAGGCTCACTGTGTCACAGAGGTTA-----CTTAGGGCTTAAATTTCAAGGACCC	474		
Db	2113	AAGGCTCAGCTCTGCTCACAGAGGTTAAAGTCTTAGGGCTTAAATTTATCCAAAGTCACC	2172		
Qy	475	AGGGCCCTCAGTGAAGAACATCCAG	501		
Db	2173	AGGGCCCTCAGTGAAGAACATCCAG	2199		
RESULT 15					
ID	AAx25663				
XX	AAx25663	standard; cDNA to mRNA; 3372 BP.			
AC	AAx25663;				
XX	21-MAY-1999	(first entry)			
DT					
XX					
DE	Human endogenous retrovirus W clone cl.P15T.				
XX					
OS	Human endogenous retrovirus.				
XX					
XX	W09902696-A1.				
FN					
XX	21-JAN-1999.				
PD					
XX	06-JUL-1998;	98WO-FR01442.			
PF					
XX	07-JUL-1997;	97FR-0008815.			
PR					
XX	(INMR) BIO MERIEUX.				
PA					
XX	Beeseme F, Blond JL, Bouton O, Mallet F, Mandrand B;				
PI					
XX	WPI; 1999-120897/10.				
DR					
XX					
PT	New nucleic acid sequences from human endogenous retrovirus-W -				
PT	expressed exclusively in placenta and useful in diagnosis and				
PT	therapy of autoimmune disease, and abnormal or failed pregnancy				
XX					
XS	Claim 1; Page 64-67; 106pp; French.				
XX					
CC	This sequence represents clone cl.P15T of the human endogenous retrovirus				
CC	(HERV) W genome. The nucleic acids, their fragments or peptides encoded				
CC	by them are markers of autoimmune disease (e.g. multiple sclerosis,				
CC	rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-				
CC	dependent diabetes and related pathologies) and of abnormal or				
CC	unsuccessful pregnancy and can be used as chromosomal markers for				
CC	susceptibility to these conditions, or proximity markers of genes				
CC	associated with this susceptibility.				
XX					
SQ	Sequence 3372 BP; 1047 A; 835 C; 711 G; 779 T; 0 other;				
Query Match 85.5%; Score 428.4; DB 20; Length 3372;					
Best Local Similarity 99.8%; Pred.No. 3.3e-128;					
Matches 429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
Qy	72	TTTCCCAATACCAAGAGAGAGAGTGGTTTACAGTCTTGAACCTTCAGAGATGCTTCTT	131		
Db	4	TTTCCCAATACCAAGAGAGAGAGTGGTTTACAGTCTTGAACCTTCAGAGATGCTTCTT	63		

QY 132 CTGCATCCCTGTACATCCTGACTCTCAATTCCTGTTGCTTTGAAGATACCTTCAAAACC 191
|||
Db 64 CTGCATCCCTGTACATCCTGACTCTCAATTCCTGTTGCTTTGAAGATACCTTCAAAACC 123
|||
QY 192 AACATCTCACTCACTGAGACTATTTTACCCCAAGGTTCAAGGATAGTCCCACTATT 251
|||
Db 124 AGCATCTCACTCACTGAGACTATTTTACCCCAAGGTTCAAGGATAGTCCCACTATT 183
|||
QY 252 TGGCAGGCAATTAGCCCAAGACTTGAAGCAATCTCATACCTGGACACTTGTCCCTGGT 311
|||
Db 184 TGGCAGGCAATTAGCCCAAGACTTGAAGCAATCTCATACCTGGACACTTGTCCCTGGT 243
|||
QY 312 AGTGGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGTGCATCAGCCACCCAG 371
|||
Db 244 AGTGGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGTGCATCAGCCACCCAG 303
|||
QY 372 CGCTCTTCAATTTCTGCTACCTGTGCTACATGCTTCCAAACCAAGGCTCACTCT 431
|||
Db 304 CGCTCTTCAATTTCTGCTACCTGTGCTACATGCTTCCAAACCAAGGCTCACTCT 363
|||
QY 432 GCTCACAGCAGTTACTTAGGGCTTAAATTATCCAAAGGACCAAGGGCCCTCAGTAGGA 491
|||
Db 364 GCTCACAGCAGTTACTTAGGGCTTAAATTATCCAAAGGACCAAGGGCCCTCAGTAGGA 423
|||
QY 492 ACACATCCAG 501
|||
Db 424 ACACATCCAG 433
|||

Search completed: April 17, 2003, 05:07:04
Job time : 209.787 secs